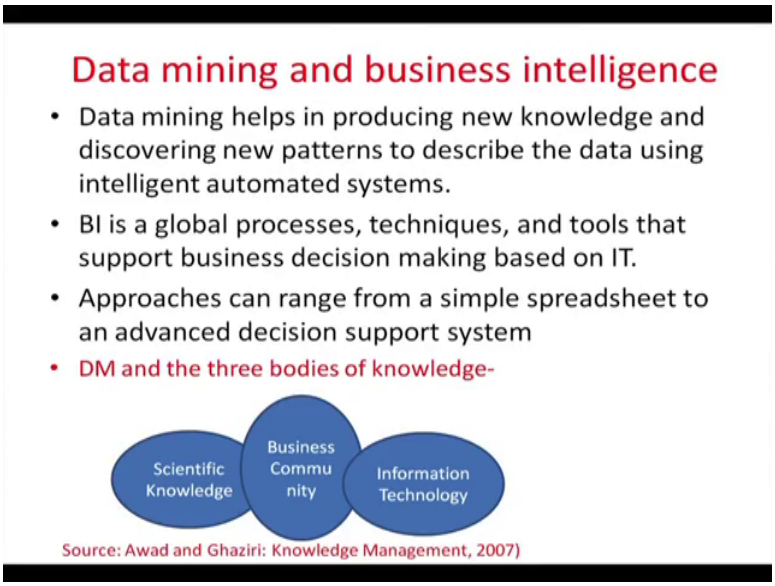
Knowledge Management
Prof. K B L Srivastava
Department of Humanities and Social Science
Indian Institute of Technology - Kharagpur

Lecture – 29
Data Management

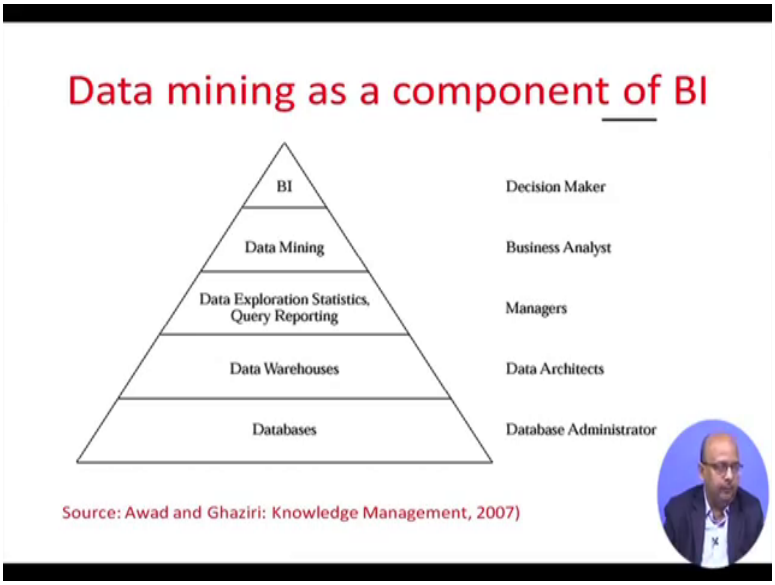
Okay, so now we are going to talk about how we are going to use data mining in the context of business analysis.

(Refer Slide Time: 00:24)



So we will continue our discussion further.

(Refer Slide Time: 00:30)

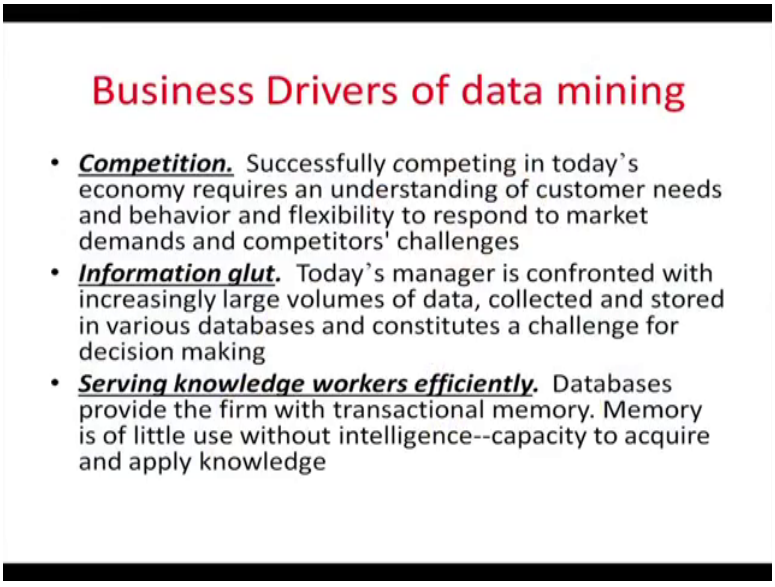


Now if you look at this it has a pyramid right. So at the lowest level of the pyramid you have Databases Okay. That is related to administration that how you are going to administer the databases. Then you have Data warehouses basically the architecture data. How the data has been kept in the warehouse okay. Then you go for Exploration Statistics, Query Reporting which is basically done by the managers.

Then you go for data mining and that is the basically the task of business managers. At the top-level you have business intelligence which actually helps you take certain business decisions and that is done by the top management. So now it creates a hierarchy where you have data architecture and data base administration activities is taken at the junior level and in mid-level managers basically there engaged in exploration statistics when they try to answer certain queries based on the data.

And then you move further so you have data mining and business and intelligence which is basically taking analyzing the data and then identifying trends and patterns and finally using these trends and patterns happens at the topmost levels where you are going to take certain decisions. So that is how it comes up.

(Refer Slide Time: 01:47)



Now look at some of the factors which are responsible for Data mining, competition, I have already talked about it okay, So since you are going to work in a very bad a knowledge intensive economy and then you had to identify your customers preferences there trends in terms of their taste we have there attitude in this kind of things and at the same time also need

to respond to the market requirements and how you are going to face the challenges that is imposed by the computers sorry competitors.

So competition because of this globalization has changed the way the businesses are being done. So what you need to do is that you need to identify that How you are going to face the challenges of the competition. So that you can better respond to the customers to offer better quality of products and services and you can also respond to the challenge of the competitors in a more effective manner right. And that is where data mining is going to help you.

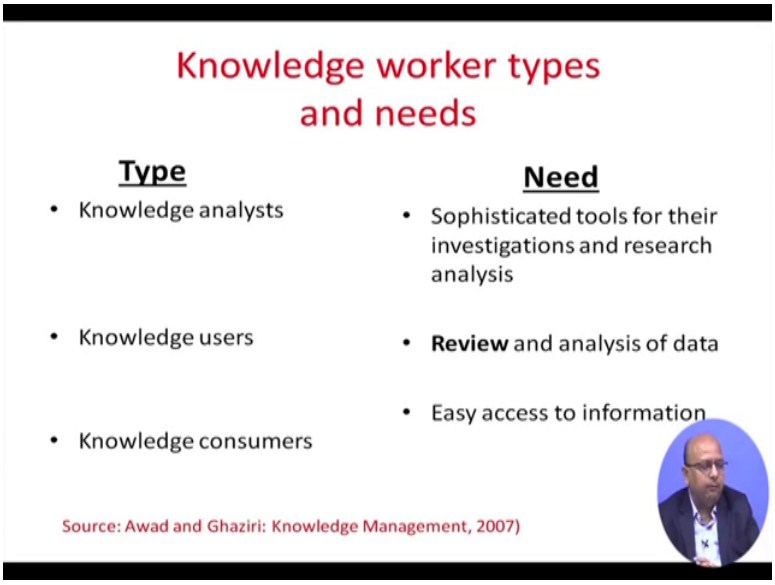
Data mining is going to help you to identify the trends and patterns which could be useful for managers to take certain business decision right. Similarly information glut what is information glut, if you are having too much information that then you are not able to make it right. So the idea is that you use start using those data that is already stored in the database okay. So you do not create a situation where you are struck with certain things and you are not able to take certain decisions based on the data right.

Instead of used databases it is not going to help you to take decisions unless you go and extract the data using certain techniques and that is why it is very important to do not allow the data to be kept idol okay. You try to make use of it as much as possible using certain analysis techniques it could be statistical techniques or even judgmental techniques. To make use of the data to say that what kind of trends and patterns are coming out and then using that trends and patterns to make certain decisions right.

So whatever data has come up okay, you have to see that how you are going to make this useful for that people were going to work in a very competitive environment may find this useful right, okay. So it is very important to not only acquire the knowledge that is coming out of the databases to that data mining but also make use of that knowledge efficiently to improve the process a system is structured so that you are going to be more competitive.

So they are the business drivers for the back-to-back certain business noticed we are already talked about it, what are the commercial benefits of this right.

(Refer Slide Time: 04:41)

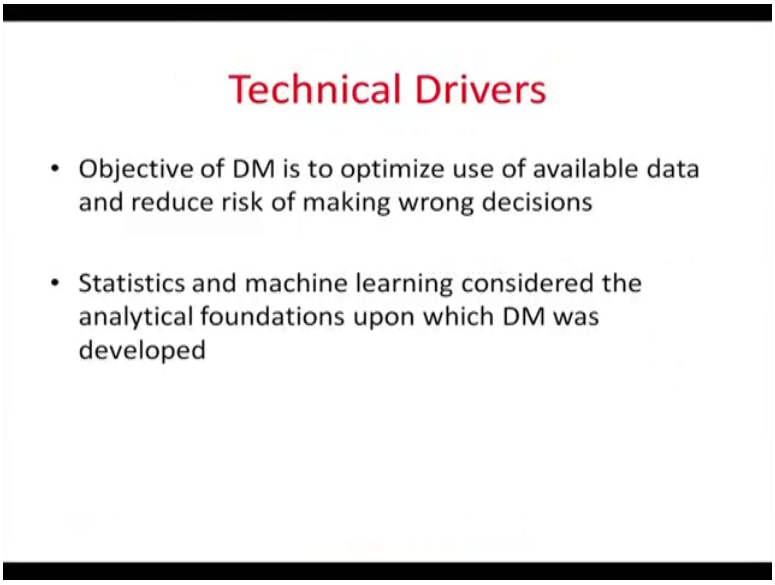


So what kind of knowledge workers are there that we are going to have today because our on the today's workplace has changed a lot and then you have to see that their requirements the there they are going to work because they live in IT information is right. Where information explosion is there having lot of communication. So the requirement may be different right for example if you look at knowledge analysis right.

They require sophisticated tools and techniques to analyze the data right. So that they can Identify the trends what knowledge users say require what is coming out of the data how they can make use of this in order to perform their job effectively. And whether this knowledge is easily accessible to them, so the job of the business analyst is to identify the trends and patterns, communicate it down the line to the top management.

And then allow these them to or help them to take certain decisions and then implement those decisions down the line so that people or the knowledge workers can make use of those decisions. And then you also have a better user interface so that they have easy access to this information okay because knowledge workers are basically the consumers of the knowledge right. So you have to make sure that the knowledge is available to them. So that is where you are going to create a K.M. system okay with a better user interface right.

(Refer Slide Time: 06:10)



Then moving to the next part as you have talked about the business drivers next part is related to Technical driver's right. What is that objective of data mining right, so basically the idea is to optimize the available data use right. What is that that you have to see that what is the best use of this data right? And you also need to ensure that you are not going to take certain wrong decisions based on this because if you are going to take a wrong decision which is not going to win interests of the business.

So you need to make sure that when it is concerned with the technology you have those sophisticated tools, softwares and techniques which is going to help you to take better decisions okay through which organizations can make better decisions and at the same time it helps them to improve their growth development and these kind of things. Suppose if you are using the data and if are not able to optimize it, then you are increasing the risk.

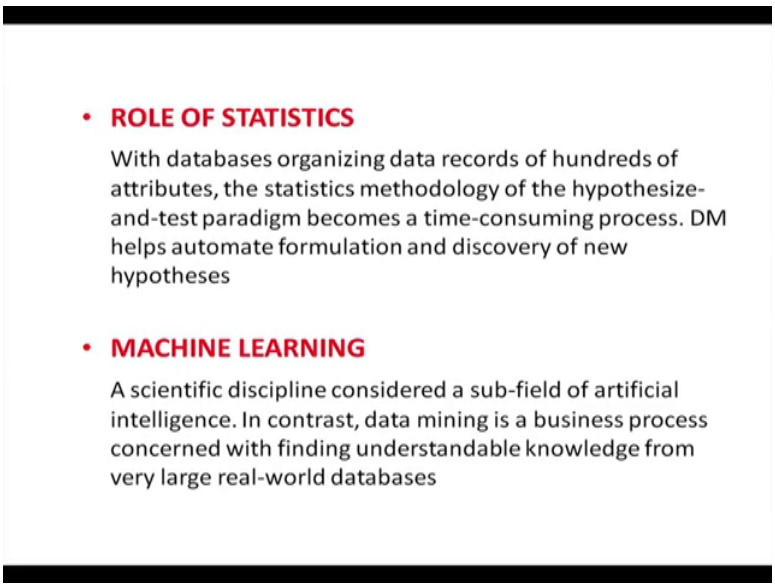
A risk of making certain decisions which may not be correct for the organization, which may not be appropriate for the organization right. So that is where we are using certain statistical techniques and also machine learning and that is basically the analytical foundations on which the data mining works right.

Though we are not going technically discuss about this but I would like to give a brief for you of these two major techniques that is used in data mining that is statistics and mission learning right. What is the role of statistics? See most of us have studied some kind of statistics in our earlier days or schools like identifying the average where I mean and standard deviation in the correlation.

And when you up in the hierarchy you also talk about regression techniques and this kind of things right. So what do you use these statistical tools and techniques and in the data mining similarly nowadays you also have lot of softwares statistical software which help you to analyze the data right. The basic idea is that means you have databases these statistical tools help you to organize in a better way okay.

These statistical tools help you to go for some kind of classification what you call descriptive analysis and also the statistical tools and techniques which are more influential in nature where you go for hypothesis testing right. So that you can use these statistical tools and techniques and even today we have a software, which are available with us which helps you to take these analyses at a very faster level.

(Refer Slide Time: 09:05)



And then use them for decision-making processes right. So the idea is that winner will to use certain software for data mining it helps you to discovery Identify new things based on either inductive reasoning on deductive reasoning depending upon whether you go for hypothesis testing or whether you want to go for some kind of explanation. Another important area is machine learning right, that is basically related to artificially intelligence okay.

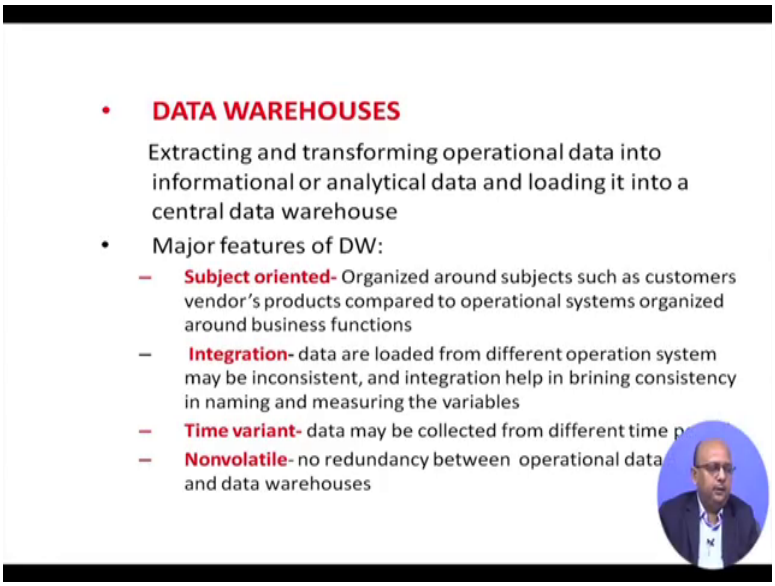
That is where you are going to feed the data into the system okay. Program them in such a way so that it is going to automate automatically analyze the data and then come out with certain decisions. So artificial intelligence is nothing else but you are going to make intelligent machines and then machine itself is going to learn how to analyze and come out with the decisions right.

So basically it is a business process are where you are going to use this these kind of a high-end computing techniques that is known as artificial intelligence to understand what kind of knowledge has been derived from the databases and how they are going to help us in taking certain decisions. So basically if you look at artificial intelligence it is the highest level of decision support system.

See we start with M.I.S that is management information system. Then we move to the next part of the decision support system that is the D.S.S. And the third level we have expert system right and that the fourth level we have artificial intelligence, it means that artificial intelligence is the highest level of decision-making process right. Where most of the job is automated by the machine and they come out with the decision.

And then you have to see whether this decision are feasible viable and whether you can go for them or not right. So both whether you are going to use statistics or whether you are going to use artificial intelligence both help you to identify the trends, patterns and values of the data and then how to take certain decisions based on this.

(Refer Slide Time: 11:15)

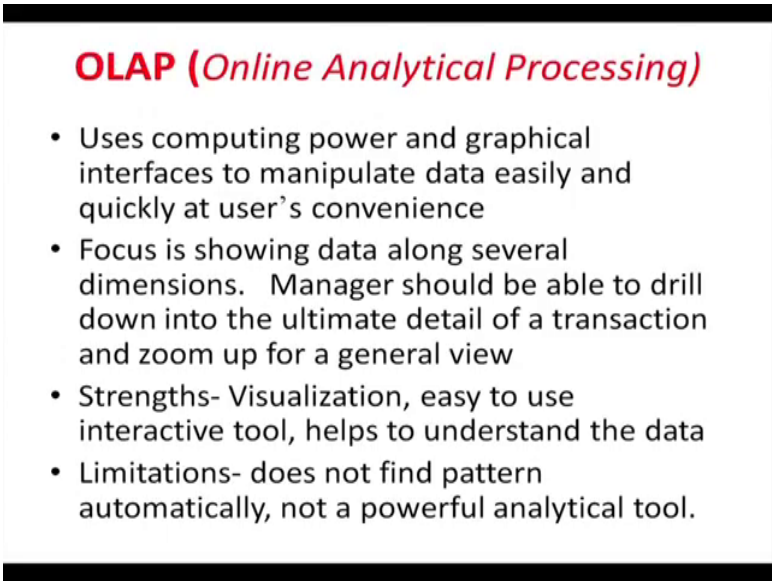


Then coming to data warehouses right so you want to extract and transform operationally data day to day data that is available with you or you want to analyze them okay. Then this is what we know as data warehouse basically it is more subject oriented for example what are the different products okay. That you have or it could be integrated also like different operating systems you have for example you have marketing system, you have production system then you can integrate different kind of data.

And you can see that our will to identify different rates and variables then time variants, is that you are collecting data from different periods which is available for example you have production data, sales data, across periods of different periods and then this data is nonvolatile. Non-volatile in the sense that there is no redundancy between operationally tousled and it did data warehouses right.

Because of the redundancy means that you must make use of it on a regular basis right, redundancy means that it is not useful it is not required but this day to day data that you are collecting the data warehouses is very, very important because based on based on that you can identify certain things out of it which could be useful for the decision-makers.

(Refer Slide Time: 12:31)



Then moving to another technique that is used in data warehousing is online analytical processing right for example you go for user interfaces you do you use our computers to analyze data based at your convenience to see what is happening right and then you can suppose you have excel sheet you can see that how the things you are moving you can use graphs also right.

So you can see that what is happening out of the data right. And that gives you generalist idea okay for example you have day-to-data of the production records right. So if you can show it on the spreadsheet, you can also make a graph right. And then analyze that how the production is working and suppose you have a target and you can see that what needs to be done in order to achieve the production targets right.

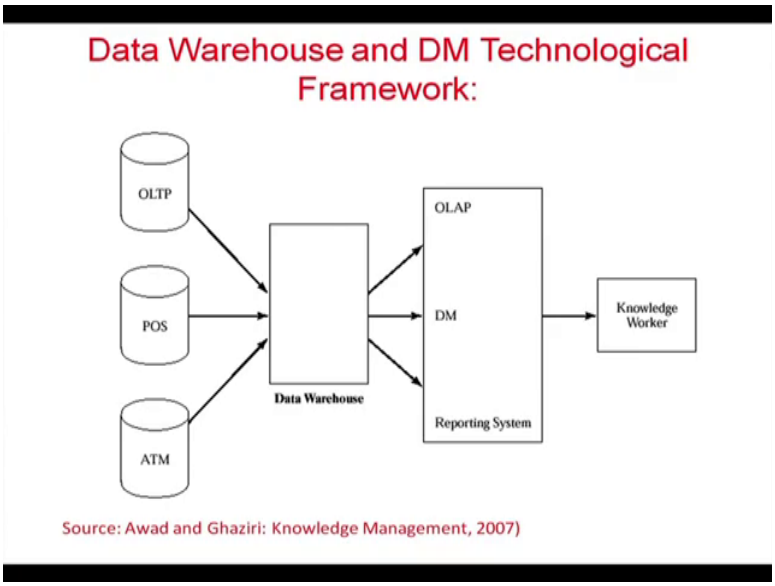
So these kinds of things are basically related to data warehousing. But it is different from data mining, the data mining what you do basically you try to identify trends and patterns based using certain either statistical techniques either it could be predictive, say descriptive or you are going to use artificial intelligence right. The strength of data warehouses visualization, where you are using certain interactive tool.

You can see that you can basically understand the data right. For example the production data and give you trend that how the production is moving, whether you will be able to meet the target production records for a particular month or not. For example you have sales data you can look at the sales data and see what is happening. You can go for a visual graphical representation to see how it is working okay.

And you can also have you can add tag certain attributes like region wise what is a sales position sales people wise what is the position, production wise what is a sales position. So that gives you a graphical visual presentation and then you can understand data. That what is happening to either sales or the production right. But it does not find any pattern, though it is a tool a good tool only helps you to provide a description so that you can understand.

What is happening right and that is where data warehousing is important.

(Refer Slide Time: 14:47)

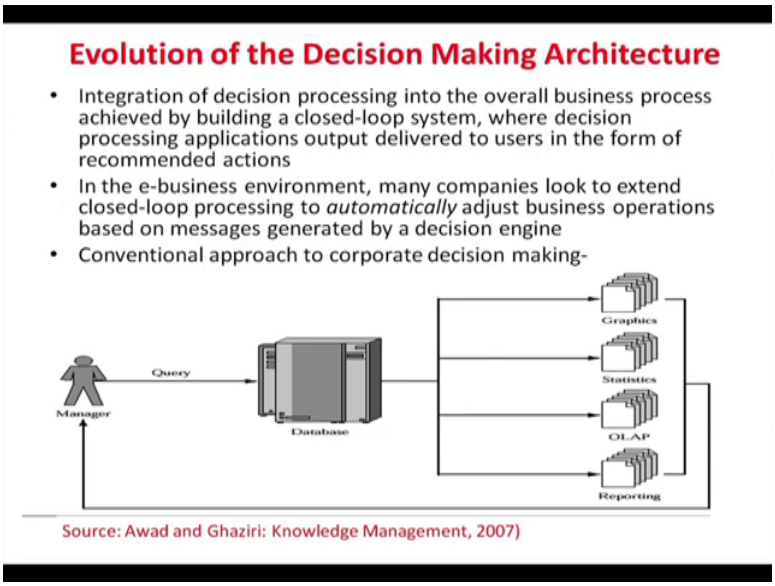


See these are the different kind of things online transactions, point of sales, ATM's all the data to the this and then you go for analyses. So you can go for three kind of things either can go for report the data as it is or you, or you go for a data mining to identify trends and

patterns or you can go for identifying processing to understand what you happen to the system right.

And then it would be communicated to the knowledge workers, so that they are better equipped to take the decisions related to their work so data warehouses and data mining technologies that frame work is different one is more productive and one is only for understanding and reporting the data right.

(Refer Slide Time: 15:32)



See how decision making architecture works okay. Suppose look at this figure and then you can understand it okay, suppose there is a manager and he has a query from the data warehouse okay. Suppose he want to know what the sales figures are across agents across products across people right. So this can be presented into different way it could be presented in a graphic manner it could be presented statistically.

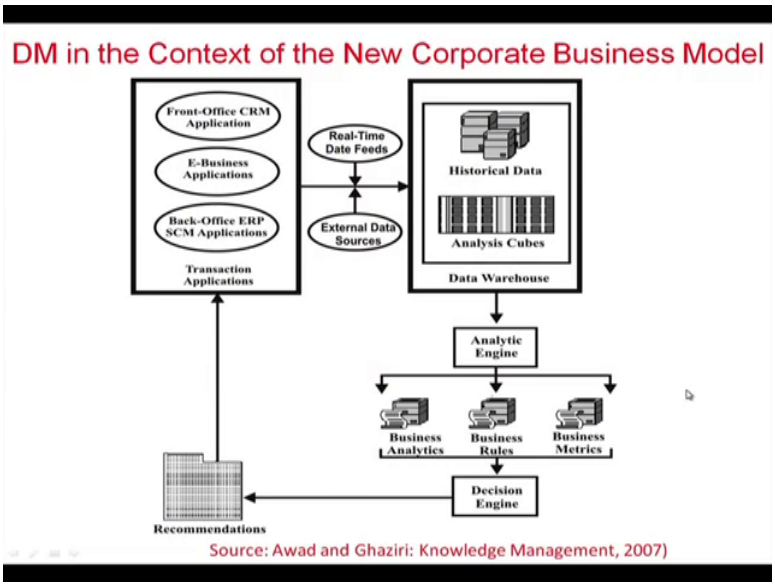
It could be also presented using online analytical processing to understand statistics see this is what we have these and this is how much is left out or you can simply report it right. So you can integrate this decision processes into a business process okay. And then you can find basically and then you are giving decision formation to the manager okay these are the thing that we have it and you can look into the business okay and what is happening okay.

So the output that is being derived from the databases are that is like it could in the graphical form or statistical form or you can use a graph or chart or simply in terms of percentage in descriptive terms or you can see rather how well we are progressing or simply reporting it

okay. So you know that most of the companies are using is a spreadsheets and data and graphs to represent or to understand their operational day-to-day issues okay.

So that they can understand what is happening and what they need to do further right. So these you can use these kind of databases and data warehouses for business operations so that you can see what needs to be done further okay. So these are the basically approaches that is used by corporate decision-makers.

(Refer Slide Time: 17:17)

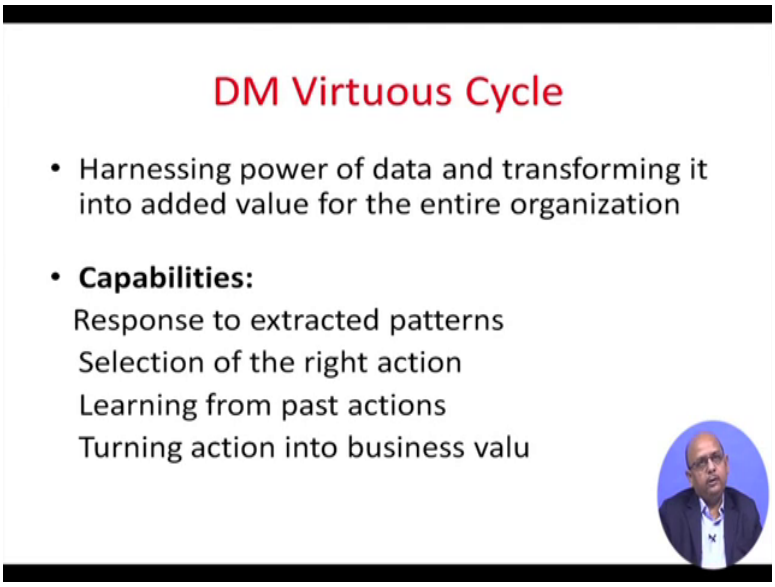


Moving further when you are talking about the data mining then it is something different right. Look at this, these are the historical data the data warehouse these are the analytical techniques okay. Either you are going to use certain metrics or rules Association rules or data mining techniques okay.

These are the and then you put into the decision engines that is the analysis and then using certain statistical tools or machine learning and then this is the decision that is coming out. Now these recommendations had certain applications it could be far ERP improvement or it could be for e-business or it could be for customer relationship management, now this is this again the feedback again has to compare right.

So the data has to be feedback to system either in real time or the data that you are getting from the external sources again need to be done it. So this is the cycle that goes on. So data mining in the context of decision-making to identify trends and patterns which are going to be helpful to take certain decisions.

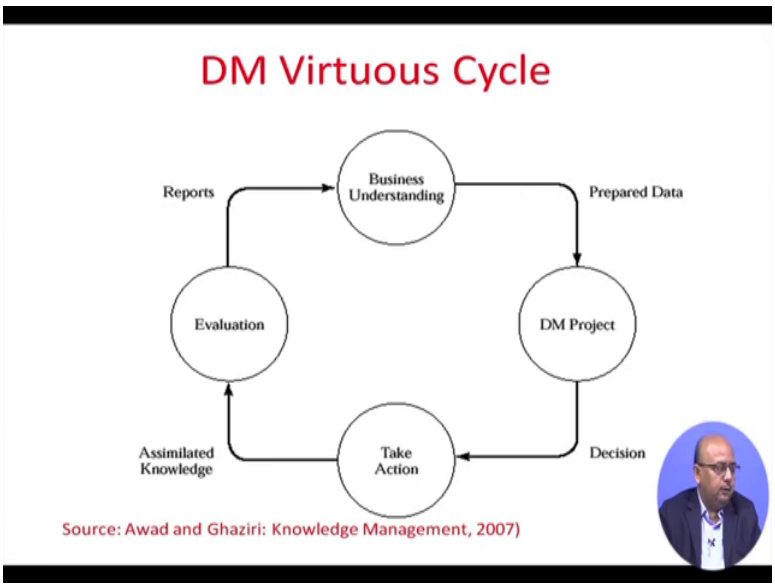
(Refer Slide Time: 18:28)



Then we have the virtuous cycle, so it is a basically a virtuous cycle which goes on a regular basis so regularly mind the data identify and extract that say patterns and trends and see what is happening to the data that is very important right. So the capabilities of data mining virtuous cycle are that you can identify patterns you can see what the right decisions that need to be taken. You can also learn from past experiences what you did in the past?

How did you make use of it? And then you can also turn action into business values right.

(Refer Slide Time: 19:01)

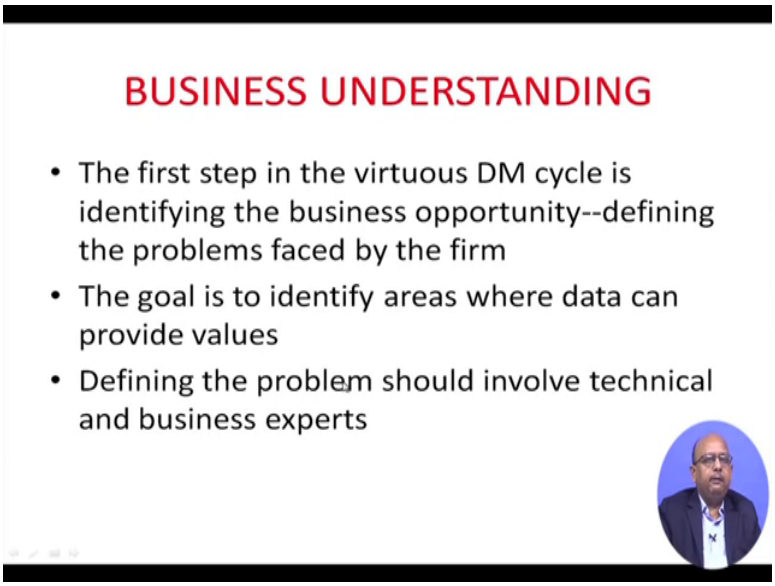


Look at this how it happens okay. So there is a data mining projects okay. Decision, action is taken then the knowledge is assimilated it is evaluated the decision is correct or not. Then it is

reported so that you have better understanding of the business this again fed back to the system or you go for it and continue with the same process.

It is a virtuous cycle of the data mining where on a regular basis you mine the data take actions evaluate your actions then see in what extent it has been able to help you to understand the business and take certain decisions and then again that information is fed back into the system and then it continues. It is a kind of virtuous cycle which continues right.

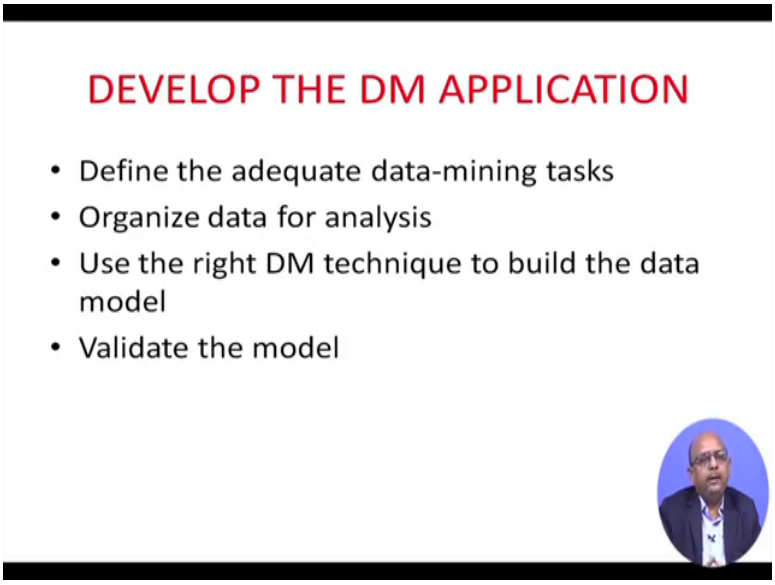
(Refer Slide Time: 19:45)



So it is very much important to ensure that data mining is going to help you to understand this right. The first step is to identify the business opportunities right. And what the problems that you are facing and in what way data mining is going to help you right. And then the next is the identifying areas where that data can provide you well. So I have given you certain ideas. So that helps you to understand that how data mining can help you in certain areas.

Like for example if you are going to mine the data related the admissions process, it can help you a lot to take certain decisions right. And once the problem is defined then you are going to see that what kind of technicalities or expert decisions are involved to take certain decisions right.

(Refer Slide Time: 20:34)

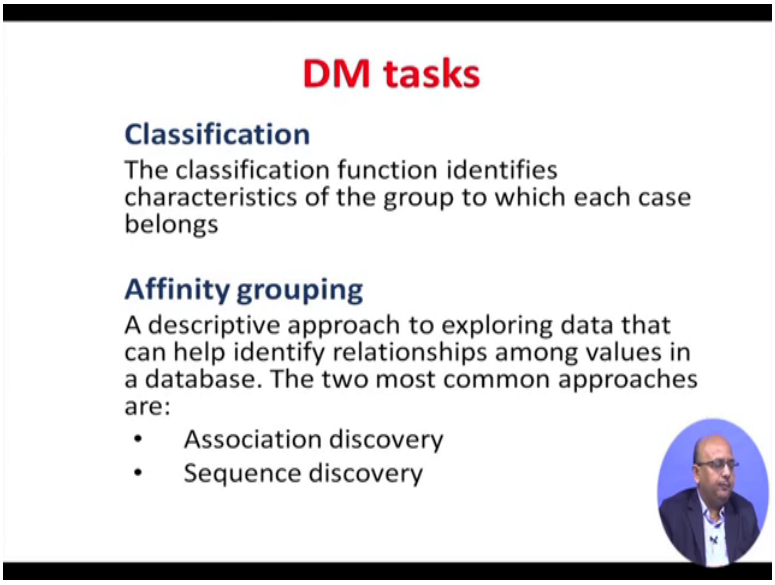


Then you are going to develop certain applications, data mining applications. What are the task that is to be done whether it is the classification, prediction or associations right? Then you are going to organize the data depending upon the requirement then you are going to be using certain statistical tools or a machine learning tasks to build certain models for protections okay.

And then you are going to validate, collect the data and feed the data into the system and see whether that data is going to be validated or not. And then if the model is validated then your hypothesis is correct you can say yes, this is related to this and that is how you are going to make use of that decision making.

So if you remember we have talked about three kind of data mining task, that is clustering, classification and association, in clustering what happens you try to find out groups that are very close to each other, but that different from others right. So you have to identify the nature and characteristics of the clusters okay. And that these nature trends and characteristic of the clusters can be tagged with that particular cluster right.

(Refer Slide Time: 21:39)



Then another important thing is classification you can classify the data into different categories and then each category will have a different data characteristics okay, this could be based on association or sequence right. We are going to try and identify relationship between different groups okay. For example you find a in JE admissions okay, from one particular state more students are qualifying with better ranks okay.

So this kind of association rules can be followed okay. Or you can also find certain frequency if this is done with this one this will happen right. So you can go for some kind of sequential analysis what will happen up one after another. Suppose if you take a political decision look at the impact of the decision right. Then how it is going to help you in the business decisions then go back to again and then you take another decision.

So this is what is known as sequential decisions based on the feedback of the first decision you go for another kind of decision right. So this kind of relationship can be identified okay based on classification okay.

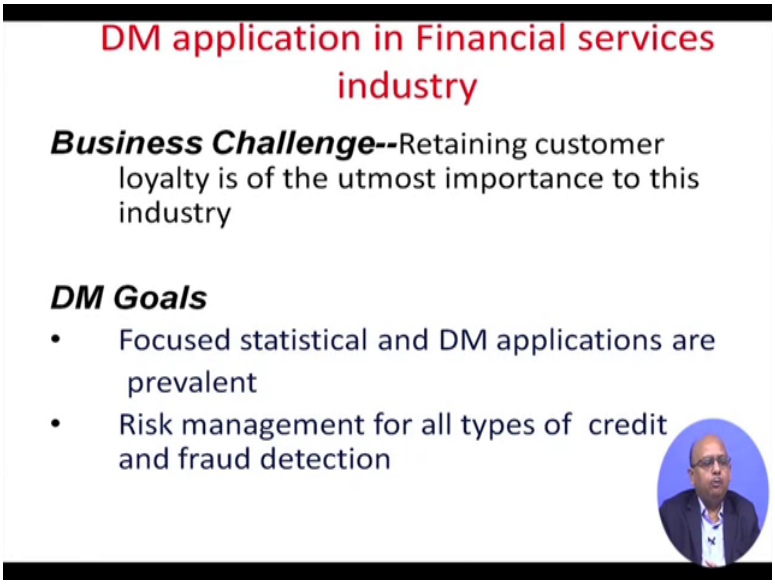
(Refer Slide Time: 22:40)

Now I am trying to give you some examples related to different areas of business that could be applied. For example in customer service is how you can apply in customer services right. What is what could be the different kind of goals in data mining okay? You can go for customers centering selling, online shopping you can identify the prediction led to the staff, what would your target market?

How you are going to retain customers who are using certain profiling techniques okay. And how you are going to use call centers were rooting your enquiries okay. So these kind of data mining goals can be achieved in customer services means that you can identify what are the preferences, attitudes and behavior of the customers so that you can better satisfy them.

With this goal you are going to use the database of the customers to mind them and then any of these goals can be identified based on that right. And then you take certain decisions relate to that.

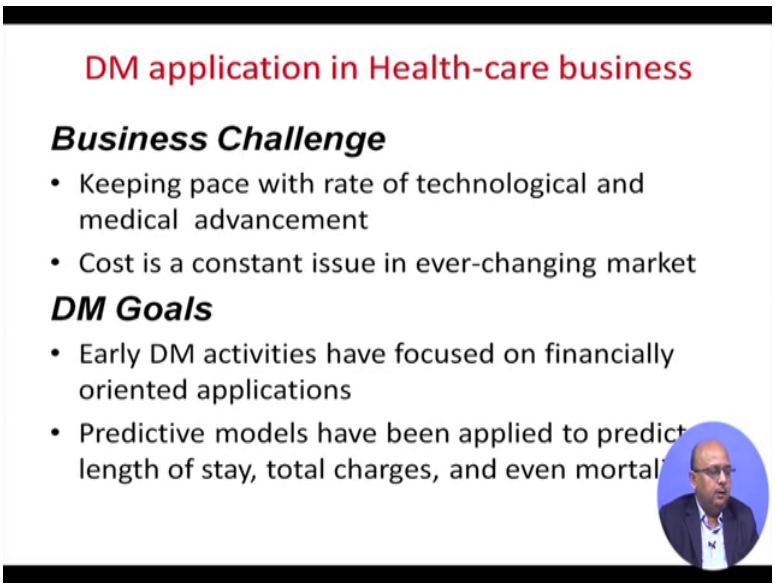
(Refer Slide Time: 23:40)



Let's take another example in financial services industry right. How you can retain customer loyalty because that is another important issue okay. So you are going to relevant to apply certain decision-making techniques sorry data mining techniques certain statistical tools to find out and see what kind of risk management is there for different kind of credit cards or how can you detect frauds.

So what you are going to develop certain applications based on that which will help you to identify customer loyalty right.

(Refer Slide Time: 24:11)



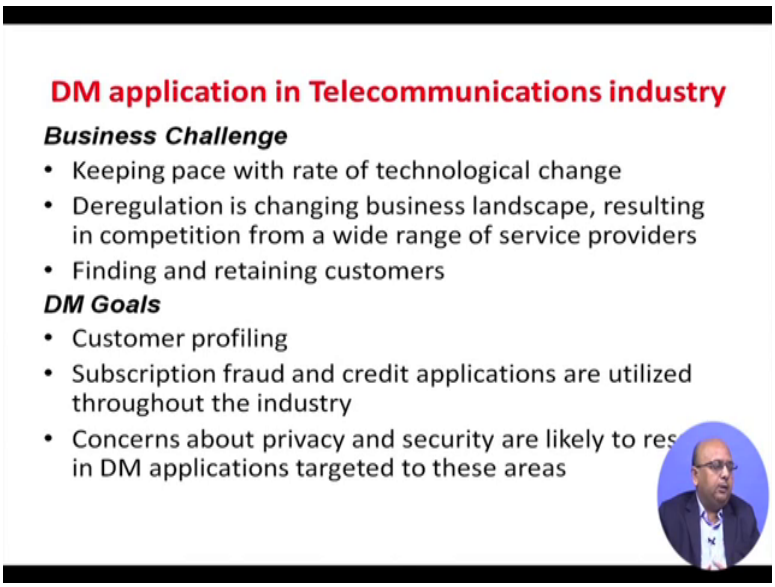
Then moving to the third part in healthcare also you can do it right. You know that technology and medical advancement is happening in a big way okay. So how can contain cast medical costs right because the market is changing and they are going work in a

competitive environment right. So not only the financial orientation will try to be cost-effective but you also see that how can, you better satisfy your customer's right.

So you can use certain productive models to identify. How long patients are staying for different kind of problems right? How much they are paying okay? What is the mortality rate okay? And this based on that you can take certain decisions that how can provide better healthcare okay.

With the constraints of the cost and on the technology change that is happening because you know the medical cost is improving like increasing like anything. Then how it is possible for you to go about it right?

(Refer Slide Time: 25:04)

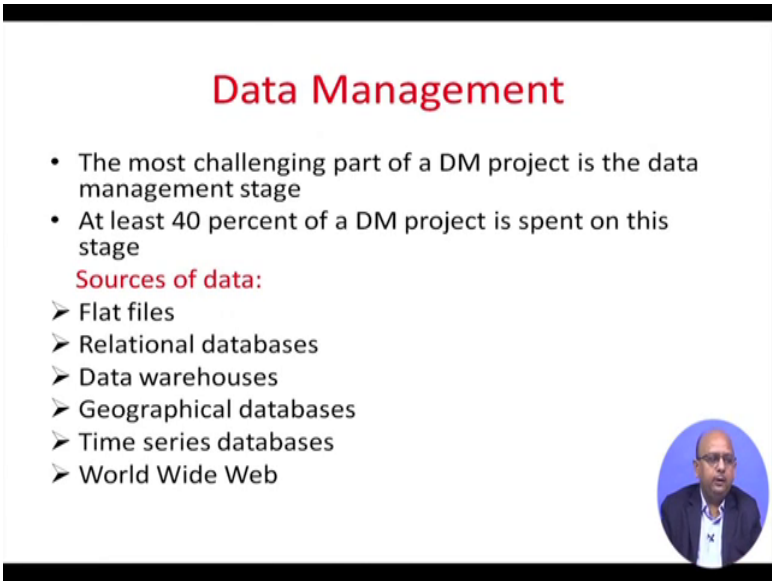


Then even telecommunication also you can identify, see telecommunication is an area where lot of changes is happening and because of the deregulation the landscape has changed right. You know if you look at the telecommunication market today. BSNL which use to have monopoly does not have that today. You have players which are effectively competing with them BSNL, the BSNL market is gone down right.

So you have to face lot of competition from a wide range of service providers. So the ultimate objective or challenges is that how you are going to retain customers? Find and retain customers. So data mining can help you to identify your customers profiling okay. You can find out what are the fraud and credit application that is used by the industry okay. You also concern with the privacy and security of the people okay.

So these are the areas which are concerned to telecom industry and that data mining activities can really help you through analysis to take certain decisions relate to these areas so that you are able to find new customers and at the same time, retain customer in order to remain competitive right. So we have talked about different kind of application in different areas.

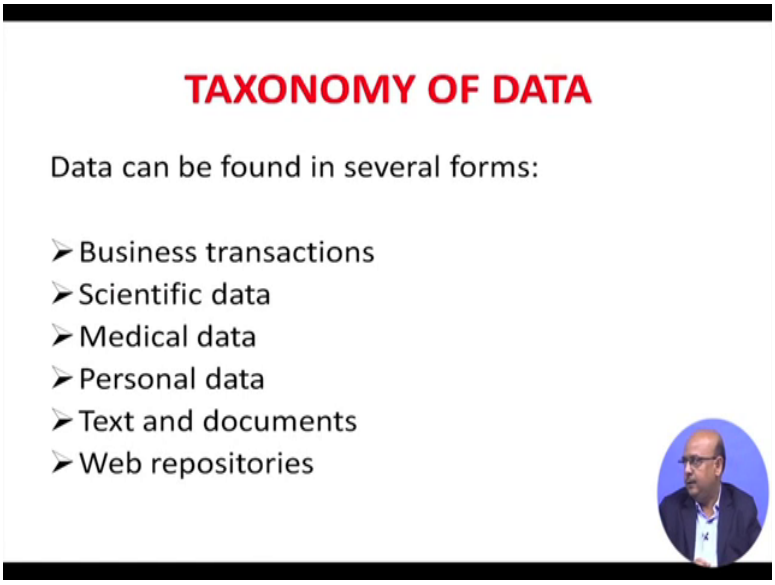
(Refer Slide Time: 26:11)



Then moving to the next part that is Data Management. Now data repository whether it is archive or warehouses whatever it is okay. It is that how you are going to manage the data. Data management is very important it could be Flat database, relational databases required data warehouses, data which bases which are spread across geographical boundaries.

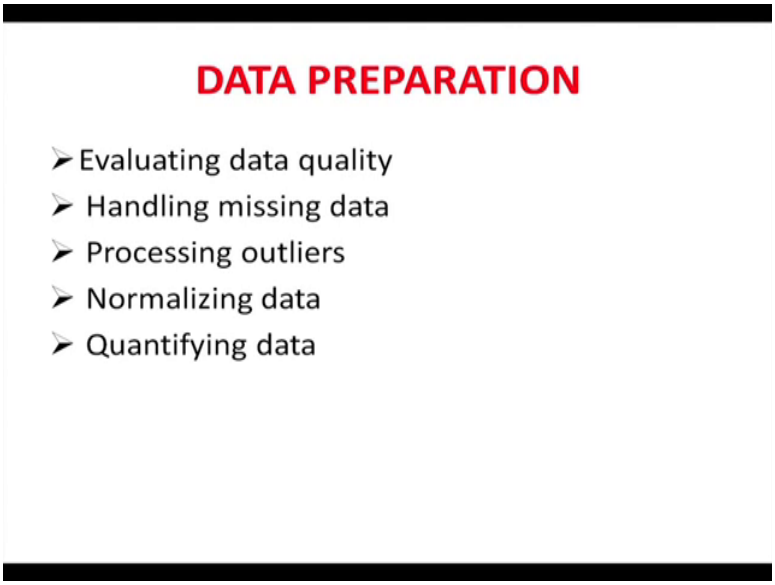
You also have data year wise data from different periods say then the data that is coming are accumulated through World Wide Web, so you have different sources of data. Now the important challenges that how you are going to manage this data right.

(Refer Slide Time: 26:51)



And then data is also found in different kind of form it could be the business based on business transactions, scientific data or medical data, or your personal employee data, Text and document data, web repositories. So you have different kind of data okay. So how when you have is the different sources and different kind of data it is very important that how you are going to manage them.

(Refer Slide Time: 27:14)



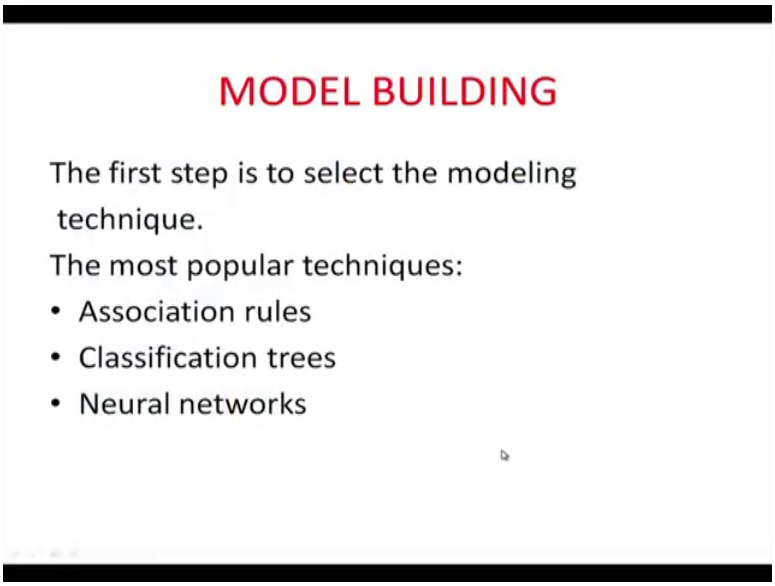
In order to prepare good-quality data it is very important to see that what is the quality of the data okay. Whether there is any missing data. What are the outliers in the data which does not conform to the norms okay? Then how you are going to normalize data? What is the process of quantifying data because not necessary that all the data that you have is quantitative in nature.

There could be a lot of qualitative data's right, like contains, blocks okay and all kind of things. So it is very important that you quantify them because if you are not going to quantify data. Say for example data related to customer satisfaction okay. So if you are used their rating scale then it could be quantified then right. So it is very important that you are going to prepare data and make sure that good quality data is there.

There is no missing data otherwise you will not be able to use it they have no outliers if outliers are there then the decision rules says that you exclude them from the analysis right. Then similarly you see that the data follow the normalization patterns okay. It falls on the normalization corps right. If it is not normally distributed it means that this data is not good and any decision based on this data may not be correct right.

And then another important thing is that you need to quantify data, whatever data is there so that you can use certain quantity techniques to advice.

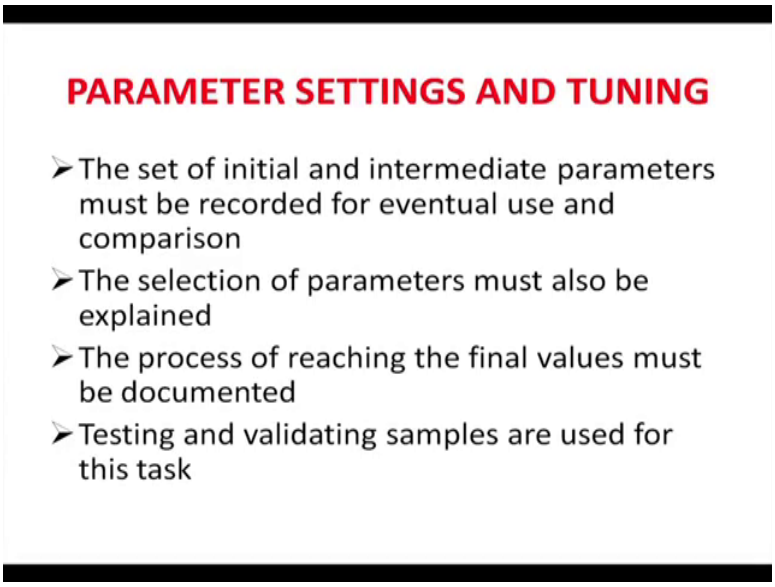
(Refer Slide Time: 28:36)



Once the data is ready then next stage you are going to build certain models okay. You are using certain rules we have already talked about it. So I am not going to discuss it here. Certain resource and rules like how one variable is used to another variables. Then you can use classification trees, we have already discuss and neural networks so these of the three modeling techniques we have already discussed in the previous lecture.

And learning from the data that how we can use these things to take certain decisions and model the data right.

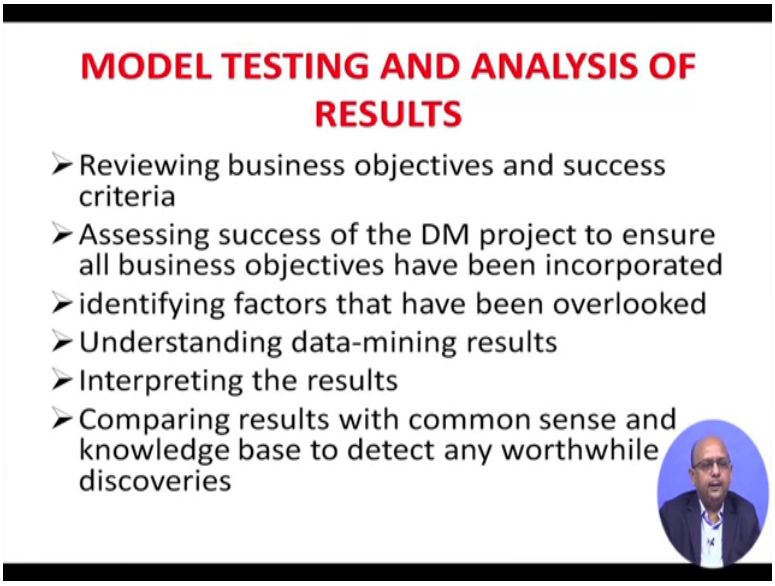
(Refer Slide Time: 29:07)



Then we are going for Parameter Setting that is you have to identify what are the different parameters okay. For use and comparisons then you need to explain these parameters very clearly based on which you are going to evaluate the data okay. And ultimately find out the value from the date right. And then ultimately you go for testing and validating samples.

For example you need to calculate reliability validity of the data which is very important because of reliable and valid data is not valid it is not providing consistent result in that case what will happen you will not be able to make use of it because it will not be correct also.

(Refer Slide Time: 29:47)

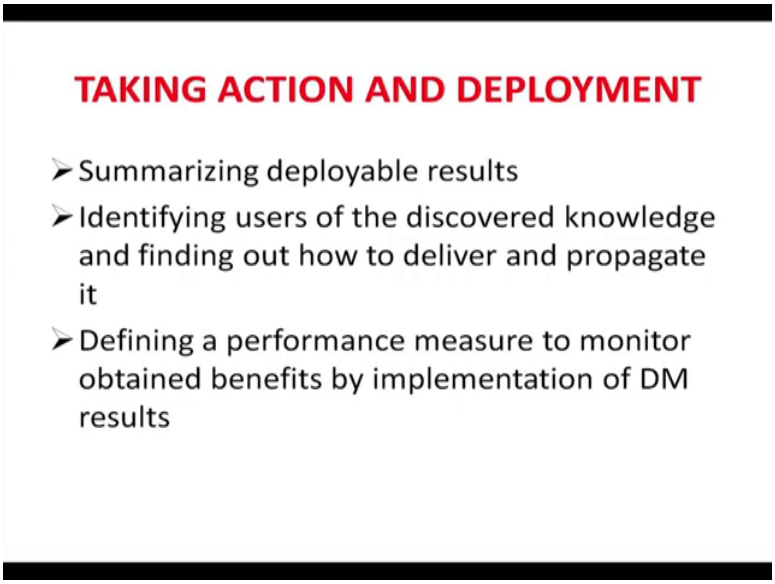


Then ultimately you go far model testing and analyzing the data right. Based on your business objectives look at your data mining project make sure that it is able to meet those

business objectives. If it is meeting it means yes you data mining project is successful. And then you identify what are the factors which have not been able to consider or have been overlooked okay.

And then once the data mining results have come out you need to understand right so the business unless you are responsible for data mining activities they must understand and interpret the results to the decision-makers right. So that it makes sense to them. Otherwise there is no use of it right. So model testing analyses of result is very important.

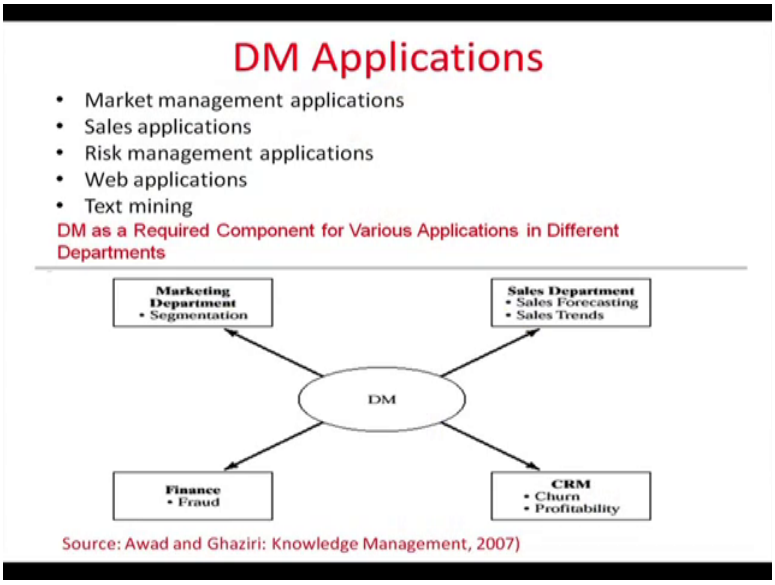
(Refer Slide Time: 30:32)



And ultimately then go for taking action and deployment, that how to you are going to take actions based on that and not enough actions you are going to take that on the basis of the results. So once you identify the knowledge then you will make sure that who are going to make use of this knowledge in workers at different levels right. And then when they are going to apply this knowledge also use certain parameters to evaluate their performance.

That whether you have been benefited by incrementing the data mining results in the organization or not right. So taking actions and deployment is equally important for data mining activities.

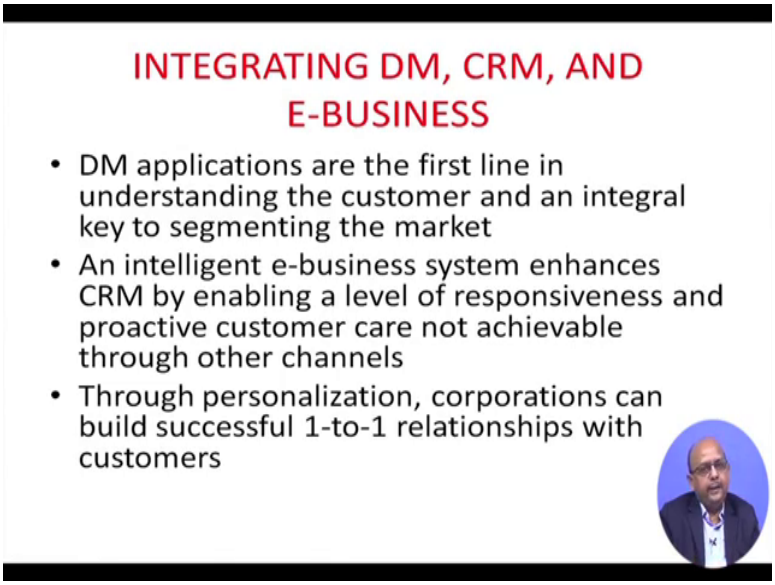
(Refer Slide Time: 31:05)



So data mining has different kind of applications it could relate to market sales risk management, Wed text mining so you can go for different kind of data mining okay. And it is very important to see that however to make use of it for example Sales you can go for forecasting and the trends or customer relationship to see that whether customers are turning or not.

How much profit were going to make of it, like finance, you can identify risk, you can identify frauds right. In marketing you can find out how the markets are segmented or how you are going to see that your advertising and sales go up right.

(Refer Slide Time: 31:43)

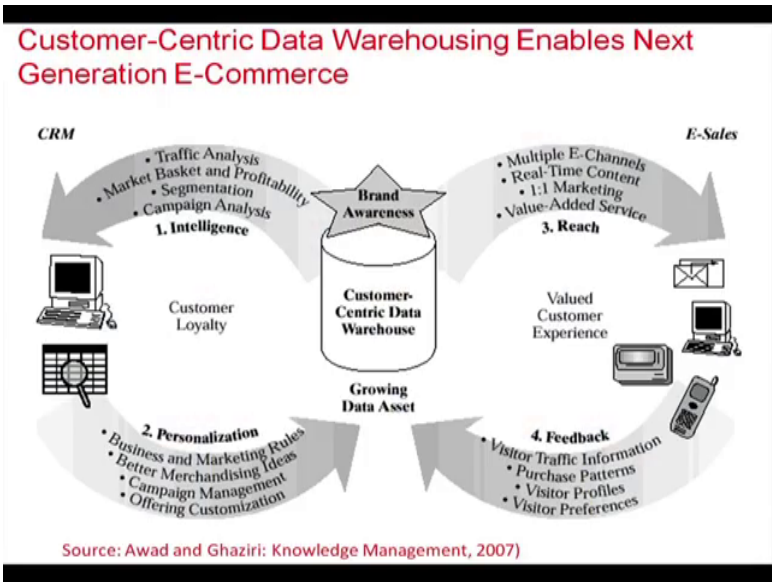


So the idea is to integrate data mining customer relationship management and e-business okay. So Data mining is the first level the next level you are moving to e-business by

enhancing your customer relationship manager so that you are more responsive and proactive so far as customers are concerned. And then you are able to achieve whatever you want to have it right.

Because in customer relationship it is more personalized approach to have one-to-one relationship right. So if you are going to integrate data mining and e-business with customer relationship probably it would be more sensible thing to do.

(Refer Slide Time: 32:19)

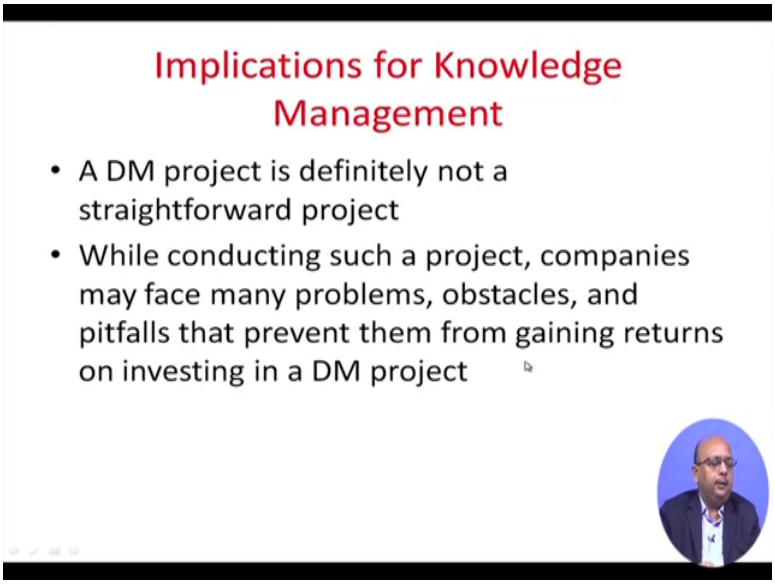


Look at this what happens in customer relationship the idea is to go for customer loyalty it could be done through analysis and personalization techniques right. Then moving to brand awareness you have customer centric data warehouses okay. And this dataset is increasing and then formation is also getting data through results reach and feedback all this data is being fed into the system.

Now what you are going to do that you are going to make use of this to see that how you are going to take it further and make use of it for taking certain decisions. For example e- sales you have the channel's real-time, contact one-to-one marketing, value-added services that relate to reach feedback is which the staff information batches by turn visitor's profile, which references, so the feedback is given to the data warehouse.

Similarly also have data related to customer relationship and then want to integrate this both part decision-making process right.

(Refer Slide Time: 33:23)



And ultimately if data mining project is not a very straight forward because you have to do lot of activities to take it as a project right. There could become constrains, projects also and then have to see that whatever you have gained from the data mining project is going to be useful for the organization okay, Thank you.