Practitioners Course in Descriptive, Predictive and Prescriptive Analytics
Prof. Deepu Philip
Dr. Amandeep Singh Oberoi
Department of Industrial & Management Engineering
Indian Institute of Technology, Kanpur
National Institute of Technology, Jalandhar

Lecture - 02
Analytics for Decision Making Support

Good morning. Welcome to lecture second lecture of the Analytics Course- Applied Analytics Course, where we are learning analytics from a practitioners viewpoint and we are going to cover all the 3 aspects of analytics descriptive predictive and prescriptive. So, today we are going to talk a little bit more on the decision making support; how analytics is going to work on the decision making and support the decision making process.

(Refer Slide Time: 00:42)

So, as we mentioned earlier in the class or in the first lecture introductory lecture in early 90's and 2000's the data mining and computational statistics these two which we can call it as the, you know they can be called as the precursors of analytics ok.

So, both data mining and computational statistics were exclusive this was exclusively used to 2 big companies because why the answer is the because of the large financial investments because of the large financial investments that are required this was only exclusively meant for their large big companies. And the smaller companies for and or maybe for other

companies who were not that big they had major roadblocks or major issues. And what were some of those major issues? The most important one would be the data collection. In those time or in the even presently also to decide you know how frequently you need to collect the data that is one question you know, what all data to collect that is another issue, then we have another one is how to store sometimes this in the form of database or file etcetera.

So, there are quite a lot of issues that comes with the data collection aspect of it. Then the second aspect of it would be you know trying to make sense out of the stored data. The question is how to make sense, and what is actually the sense that aspect. The third aspect is decide what all techniques to use; techniques to be used.

So, these were major roadblocks because these questions are usually they or differ from organization to organization; they all differ from organization to organization and within an organization problem to problem. So, the requirements in this case is quite difficult. And, so then what happens is analytics slowly matured from what we call as this business intelligence. So, the organizations initially who were doing data mining and computational statistics that require the first step is first step is collection and storage of data of let us call it as relevant data which is what we call as the business intelligence. And once you have that, once you have the data being stored it actually slowly matures and to analytics. So, what happens is BI involves collection of data and information for future usage that is one part. Then, once the data plus info available opportunity for predictions estimates etcetera; you can see predictions and estimates slowly starts to come into picture.

So, then this may BI; so if you think about it does time progressing the time is progressing in this regard matures to analytics. So, when you have all the data stored and then you have the data available with you then you start using that for doing analytics. Hence, plus the advent of open source tools drastically reduced cost of analytics. For example, SAS versus R; SAS is a very expensive software which is used for earlier statistical computational statistics and data mining, but now there is an open source version of it which is called as the R ok.

So, this all resulted in from niche organizations to look small and mid sized organizations. So, the analytics started being from the big companies to the small companies. So, if you think about it in a graphical format as the time has progressed. So, here we had only those big corporations it is kind of drawing big corporations like this they were only doing what I call us data mining and stats computational stats this was being only being them by them and the

finance money was like lot of money was involved in this as time progressed what happens is one was new tools new free tools then you heard other thing was called cheap storage and then you had cheap hardware cheap power computational power etcetera they were all coming in then you hard like all the medium sized organizations and the small sized organizations reasonably doing analytics.

So, then the BI translated to analytics that is how it actually evolved.

(Refer Slide Time: 09:11)
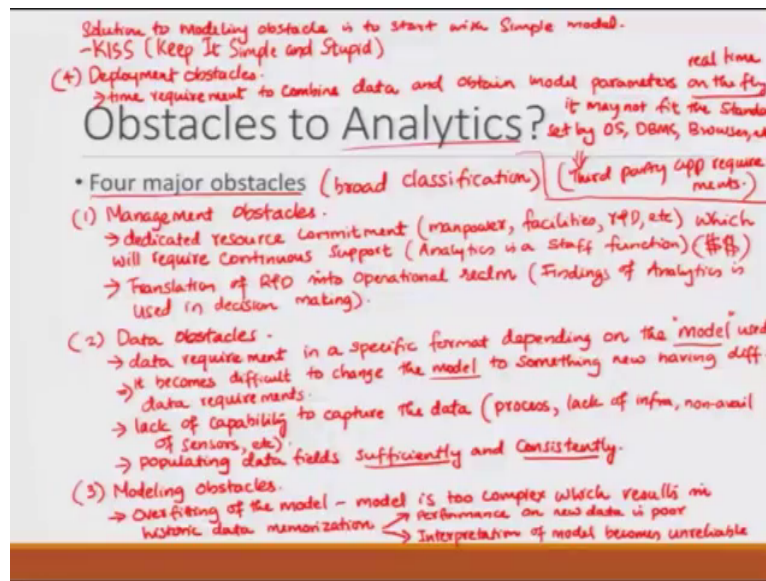
So, now we will talk to about the next ones what are the major obstacles to doing analytics in an organizations there are 4 major obstacles you can there are there could be many, but we can classify it into 4 major. So, this is a broad classification you could find any of these in any organization and if these are fine identified then it need to be addressed. So, the first one is called as the management obstacles is a category. So, what are the major management of circles the most important thing that you would see is the dedicated resource commitment the resource in the could be in the form of this could be manpower then you have facilities then R and D etcetera. So, these are all dedicated commitments which will require continuous support.

So, because in organizations mostly analytics is a staff function a dedicated team is usually available for doing analytics hence you require dedicated resource commitments which means you how to invest money in this regard financial investments then second part is translation of R and D into operational realm. So, what happens here is findings of analytics

is used in decision making though this sounds simple, but many a times this is actually an issue in industry purely because of the fact that the decision maker the large scale decision maker or the big decision maker usually have a tendency not to follow the recommendations of the analytics team they are like I know better I have been this job for. So, many years why should I listen to a bunch of people who are not made any big decisions.

So, the because lot of people consider analytics are still as an R and D input and they there are there are many incidents where are these such inputs are being discarded and finally, if you take in the long time you will see that the analytics will be give telling something and the decisions it will be going in the opposite direction. So, that should be avoided. So, for that the management need to put this food down and basically get the things done asks and say that well we are going to do data driven decisions. So, and for that we are spending this much of money. So, better we should show outcomes for it. So, that is the one part second one in this regard or second obstacle in this regard is the data obstacles.

So, this is common in many many ways in many organizations and data requirement in a specific format that is usually the biggest problem specific format depending on the model used. So, what we are trying to do say here is that depending on which model you are going to use the data that you are going to do would require in a specific format. So, usually what happens is. So, it becomes difficult to change the model to something new having different data requirements. So, you would probably see that many other organizations that actually do analytics they the changing of a model to a new advanced one with a very different data requirement they tend to avoid that because their data storage the business intelligence require a large amount of revamping which requires a lot of money in that regard the other time is lack of capability to capture the data it could be due to the process like a high temperature process or something like that or it could be due to lack of infra or known availability of sensors let us say you are taking some data etcetera.

So, sometimes you cannot capture the data just like that like say for example, if you look at if you do not have the sensor to do a particular temperature or select. For example, you want to do some process in which the temperatures are extremely high and you do not have a sensor that could withstand the temperature then you cannot capture the data. So, that is that is the point. So, sometimes the process will make it extremely difficult to capture the data sometimes the other factors would make it extremely difficult to capture the data and

sometimes it becomes such an issue that you cannot capture the data that frequency that is needed.

So, it becomes an issue. So, then finally, the last part is populating data fields sufficiently and consistently. So, the data fields need to be populated sufficiently and consistently. So, you had should have consistent data and you should have sufficient data then the third point that we are going to talk about it is the modelling obstacles. So, you have data obstacles we kind of alluded to what would be a modelling obstacle would be the modelling obstacle typically is called as the major issue is over fitting of the model the model this literally means that the model is very too complex based on the data you our model is too complex which results in memorization of data in historic data memorization.

So, when it results in historic data memorization what happens is that this creates two issues the major two issues the first issue is that performance or new data is poor the other part is that interpretation of model becomes unreliable. So, the when you provide new data because it starts memorizing the historic data when you provide new data it has a poor performance and the model becomes unreliable when it is interpreting when you interpret the model it becomes an unreliable thing. So, these two are the outcomes of the modellingobstacles. So now, how do fit solve this modelling obstacle I am going to write on the top solution to modelling obstacle is to start with simple model I cannot emphasize this much more. So, we are going to use a new principal called kiss principle which means keep it simple and stupid.

So, start simple this is the big mantra that you require when you are doing modelling in especially in data analytics and then the last one the last obstacle that we are going to talk about is the deployment of obstacles and the deployment obstacles the major thing is the when you implement both the data and the models the time require the time requirement to combine data and obtain model parameters. So, when we are talking about this we are combining data and we are obtaining model parameters on the flight on the fly means real time when you are doing it in the real time process then I know when you are doing this it may not fit the standards set by operating system DBMS browser etcetera.

So, many a times we will see if you look at if you book your website train tickets in I R C T C you actually see that sometimes it takes a long time and sometimes it is take a time operation timeout because it is trying to combine data and applying the model onto the top of it, but then it actually means the session timing of beats a session timing of the browser and

then you get a session timeout. So, sometimes the other party operations like things like the third party software or third party applications requirements this is very hard to be certain times. So, that is some of the reasons why when you have a very complicated stuff that you are working on then you will see specific or custom built system made available for this process ok.

(Refer Slide Time: 21:25)

So, we now talk about the next process why do we need decision making everybody talks about decision making the most important question is why do we need to make decisions well the first and foremost thing is rational decision making helps the organization it helps the organization in many ways. So, the first and foremost thing is the decisions when you are taking better decisions when somebody takes a better decisions it results in it results in focused actions better decisions translates to focused actions your actions are focused and reactions are focused you get close to desired results.

So, the reason is that better decisions also translates to clearer goals. So, if you are taking better decision rational decisions or data driven decisions your goals why are you making the decision that is very clear and once you know why are you making the decision then it will give you a give right to focused actions once the focused actions are there then you get very close to the desired results and when you decide results mostly in an industry it is profit and efficiency.

So, you increase your profit you increase your efficiency then there is something what we call as a scientific decision making approach which is where you know there is another one other alternative to the decision decision is not the emotional or you know we can call it as whimsical decision making is not is like not my hunch I am just not I am going to make the decision because I feel. So, not that way it is a scientific decision making approach it is based on data and there is a specific process involved and it will discuss this in the next slide, but what is it does is it reduces the time reduces the time means time to take a decision this is the classic example of this is bureaucracy the time taken for them to make a decision is extremely large and hence it translates to what you call as risk.

So, as time progresses the risk increases this is this need to be thought through and then; obviously, cost also. So, if you decide that we are going to build a building today it might cost you one crore rupees, but if you say that no we will wait we will wait for 2 more years and the mule builds a building it would probably then cost double or triple that time money. So, that is an example of by delaying the decisions are waiting for more time usually it does not reduce the risk and the cost it actually do the counter productivities usually increases it and the other part is that rational decisions when you talk about rational decision making it is continuous in nature this continuous in nature means you take decisions and you follow up there is follow up which means it results in a closed loop system.

So, if you think about it you have the inputs of the decisions and you have the system you make your decisions and you have your outputs and typically if you are making rational decisions there is a feedback always and the system will slowly you will tend to what you call as a course correction it will correct itself. So, that the expected behaviour of the system becomes what is it desired. So, you will reach the desired behaviour in as time progresses.

So, time is kind of continuing here then the another part is it also empowers the people to make decisions what is what is the major important part of it, because if the decision maker is not empowered the person who is supposed to make the decision is not very sure that if I takes a take this decision. And this is the decision that are taken in my good faith which means based on all the data all my emotions were thrown out of it or taken the decision properly and I hope that the organization the upper management will stand by me.

So, there if you know the upper management is also making rational decisions they only look at the facts and they do not look at other things at all if that is the case then everybody will be

happy to make the rational decisions they will not think whether if I make the decision then will that person be happy or this person be happy that kind of if that kind of questions are eliminated whether the happiness or individual feelings or emotions are taken away from the decisions and you just focus it by using data driven decisions which is what analytics do then rational decisions people are empowered to make decisions the final point is it also increases the organizational intelligence. So, organizational intelligence is increased by how is it increased it is increased by being informed controlled and adaptive.

So, when we say organization intelligence what do you mean by organizational intelligence as people learn organization is consisting of people as people learn to, but take decisions what happens or when you empower people to take decisions they tend to act they act more intelligently. When people act more intelligently what does it do it translates translate to organization acting intelligently. So, the organization acting intelligently literally means that the people in the organizations are acting intelligently and their intelligence actually come from being empowered to take data driven decisions for which they happily use analytics. So, you can kind of see how the analytics and can help in this critical process called decision making now let us.

(Refer Slide Time: 28:18)

Talk about the decision maker the person who talks about who makes the decision what are the major factors that influences the decision maker and the decision maker is a person who

exhibit a range of beliefs and the belief in this case is not the religious belief or anything, but what we talk about here is how strongly one accept or differs on issues.

So, we talked about is here is the degree of acceptance or difference on an a issue and what are the major factors influencing this belief or this individual act acceptance or differences on issues the number one is the evidence well this is not the evidence of the law enforcement, but this evidence is called as what we call as the what is the prior knowledge what is the prior knowledge in this regard knowledge about the issue about the issue which means it can in a better way to say this about is domain knowledge domain knowledge. So, do you what do you know about what is already going on. So, sometimes you know more the domain knowledge more the domain knowledge it can translate to stronger the belief stronger beliefs ok.

So, such one part dividends second part is the compatibility. So, here a when we are talking about compatibility that we are talking about the agreement with the worldview of the individual how compatibilities like; so like a for example, of this is let me put us an example here the trickle down effect of money. So, what happens is no pay top people more people at lower position will also get more eventually this is a belief that a lot of people do harbour and I am not an expert to say this, but people do believe this the best way for giving money to the lower employees or something needs to pay the upper management high. So, when the salaries are high then slowly it will start trickling down to the lower employees because they will also like would like them to bring up bring them up higher because the differences will be huge and those kind of things. So, if the person believes that then the actions will always be the decisions will always be dependent upon that world view that view that the reset trickle down effect of money.

So, these kinds of things can happen then the third party is the third factor that is about his reliability how reliable is the source of information or data. So, a classic example of this is you know government provided data versus info in social media somebody goes there facebook and post something and some government official comes and says we have absolutely no idea with this is we do not agree to this; Then, obviously the people would start to believe the government source because people would say that one source is more reliable than the other kind of a thing.

So, the how reliable the source of the information of data is also one critical factor that typically influences the decision maker because if we know that the source is reliable then tend to believe the data more or you tend to your belief is better the last part is the accepting and doubting here when we talk about it is what are the consequences of accepting or doubting the decision. So, for example, the main example here is voicing difference in opinion.

So, if you know that, if you do not agree with the top management you are going to voice a difference different opinion and if it is going to heard you if it is going to create issue to you then; obviously, you may not be comfortable in doing it you would probably decide that fine I am going to stay away from this what I am not going to voice my opinion or difference the other hand if the organization promotes such kind of voicing of difference of opinion then; obviously, you will tend to voice it strongly and all the difference of opinions are heard another example.
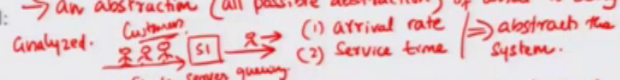
So, for the example of a company like this is Honda who actually promotes you know opinion difference. So, you should Honda always used to say that if you know the other person is an investment to you one of you is different one of you is different or obsolete kind of a thing, so anyway. So, all these factors influence the decision maker or the belief of the decision maker which then translates to the decision. So, when you provide data when you have reliable data which is compatible with the worldview of this and there is prior knowledge evidences or hypothesis available with that and the consequences of accepting or doubting are you know understandable and clearly articulated the decision maker tends to make better decisions or rational decisions now we let us talk about what is this scientific decision making process or SDM ok.

People we already mentioned there is a scientific decision making process involved and there are many versions of this available different views on this people provide different views about it, but in almost all the views you can kind of find these 5 major elements in most of the variants of these views there are 5 major elements inside and what are these 5 major elements the first one is called as the goal and the goal literally means objective of doing the analysis why is the analysis being done ok.

So, this is the first part of it goal second part is the model it is typically an abstraction in many cases it is all abstractions all possible abstractions abstraction literally means you are already taking what is needed and unwanted things are thrown out all abstractions of what is being analysed. So, an example of a model is that let us say that there is a bank teller. So, this is a bank teller and there is a queue people come here and wait at the teller and then they get the money and they go away people go away out of the system that is the thing about a single simple system which is called as a single server queuing. So, the factors that you are looking predominantly here one will be the arrival rate the rate at which the people are arriving and the other one is called as the service rate or service time you do not really care what colour of the dress this server the teller is wearing or what is the dress the race cars creed of the customers.

So, these are your customers. So, you will only take these 11 information and the rest of the information is distracted. So, this are discarded. So, this abstracts the system. So, that is what

we talk about here as a abstraction. So, more all models are an abstraction of the system that you are going to study third part is the data. So, what is data that are observations that represent the scenario or process in which the model is to be tested? So, the data gives you the scenario or the process where the model needs to be tested the model is abstraction of what it is and then you are going to test it on a forgiven scenario which is represented by the data and then what do you have is evaluation.

So, what you do here is compare the model to data and decide the suitability of the model you are saying whether this model is working for the data the abstraction that you are using is working for the data and then revision basically means change the model if evaluation results are not. So, you are saying that fine we evaluated the model with the data and then we found that it is not fitting it is not compatible. So, hence we change the model. So, these aspects the goal these 5 steps the goal the model the data the evaluation and the revision is quite common in the scientific decision making process. So, what we do here now is we.

(Refer Slide Time: 40:01)
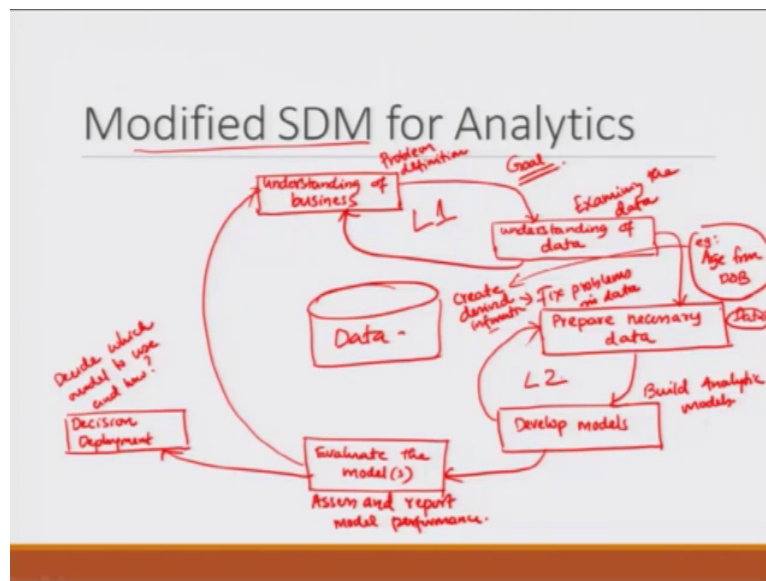
Provide a modified scientific this making process modified SDM for analytics for the analytics that we are interested in this particular course. So, we basically start with let us start with the no diagram means always you can start drawing boxes.

So, we will say understanding of business let us call that as the first starting point of box and then from there we go to something called it as understanding of data. So, from here we move here. So, once the what is you think about understanding the data sometimes what happens is

once you start playing with the data you get something called as a better understanding of the business and this loop can continue a couple times then once you reach a reasonably good understanding of the data you move into what you caller said prepare necessary data prepare necessary data. So, once you understand the data you move into what you call as a preparation of the necessary data then once you have the that part then what we do is we basically move into the next step which is called as develop models you go in this fashion sometimes when you are developing the data you find out that fine or model then you say that fine the data is not in the right format then you and then end up iterating there finally, at one point of time when you are comfortable with what the model is then you do the next step which is called as evaluate the model or evaluate the models.

So, from the development of the model you evaluate the model and once you have the, but model with you can do this it actually would better your business understanding and you could repeat this process which is another loop or once you are comfortable with this then what do you call as the you do something called as deploy the decisions or decision deployment when your model fits whatever it is then you can say fine I am using this going to use this model to deploy the stuff. So, where is data in this whole mess the data is at the center here is what the data is this whole thing is happening around the data. So, the understanding of the business is the what we can call about us the problem definition or here we can talk about it is an understanding of the data where what we are trying to do is you are examining the data alright.

So, these things translates to what we call as the goal of the aspect of this system then we do talk about it is you know we prepare the necessary data where literally means that we are fixing the problems fixing problems in data and in also one part of this is that create derived information a classic example of this would be somebody's date of birth might be given to you and the. So, so age from D or B is an example this is a derived aspect fine now once you prepare the necessary data then what you do is you develop models and. So, here what you are doing is build analytics models this is the modelling aspects of it. So, ask the earlier reason the goal model day.

So, this is the goal this is the data which is especially here then you have the models once you are the models then you have evaluation which is assess and report model performance this is the evaluation aspect of it and the decision deployment means decide which model to use and how and how are you going to use the output of the model to make decisions.

So, this is the sub s d m that is typically available or typically practiced in the industry you start with the understanding of the business then you understand the data. So, we call this as the loop one and we also call this as the loop 2 we need to talk about these feedback loops L1 and L2 and we will discuss this immediately.

(Refer Slide Time: 45:21)
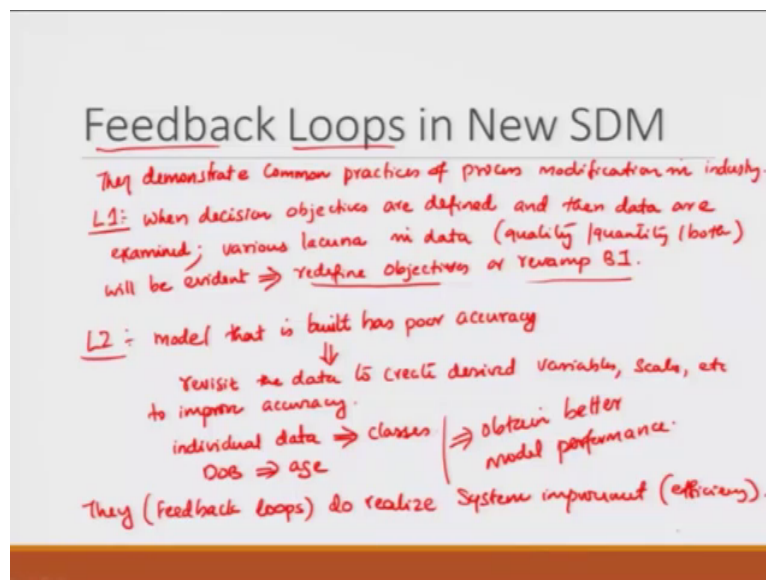
So, these feedback loops there are 2 loops. So, what do they demonstrate the feedbacks loop and they demonstrate common practices of process modification; so modification in industry.

So, what we do here is the loop 1 L1 what we talk about it is when decision objectives when decision objectives are defined and then data are examined when you are done your digital when you defined your decision objectives and then you are examining the data then various lacunae or deficiencies in data this can be quality quantity both elevens all these aspects will be part of this you know will result will be evident what you would do is redefine objectives or revamp BI. So, either you would redefine your objectives or you would change the process in which you are collecting the data.

So, this first feedback loop typically is refinement typically a process refinement that happens there itself most of the time the second loop L2 that we see in the previous slide that is based on you know what happens is that the model that is built model that is built has poor accuracy. So, then when you have that what you do you revisit the data to create derived variables scales etcetera to improve accuracy. So, sometimes you find that instead of the

individual data points. So, you can say that individual data point's data may be translated to what you call as classes date of birth translated to age etcetera like this.

So, sometimes you might come across something some derived variables which might give you a better performance to the model. So, that the internal modification of the data is done then and there. So, this is to obtain better performances of the model better model performance. So, these two feedback loops in reality do you know happen during the scientific decision making process or analytics within the industry and these two feedback loops to a large extent they improve system they which means the feedback loops do realize system improvement where efficiencies definitely improved which will translates to better financial returns.

So, I hope that at least we you understand from the way how analytics is supporting the decision making and why decision making is important and what is a scientific decision making process. The aspects behind the scientific decision making process that aspects of these course, I hope it is clear to you guys. And in the next lecture we will cover the new concepts data and other aspects what are the type of date and other things. So, till the mean time continue reading and learn more and more as you can wish you all the best.

Thank you.