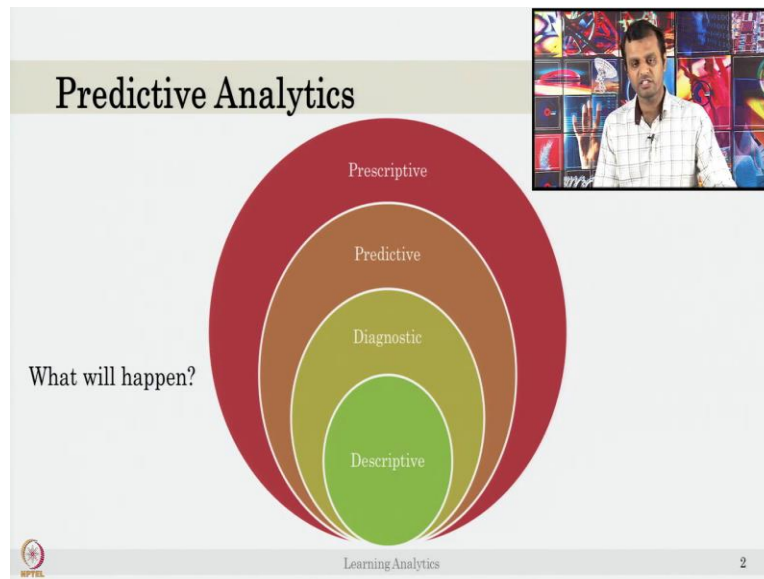**Learning Analytics Tools**
Professor Ramkumar Rajendran
Department of Educational Technology
Indian Institute of Technology, Bombay
Lecture 8.1
Predictive Analytics
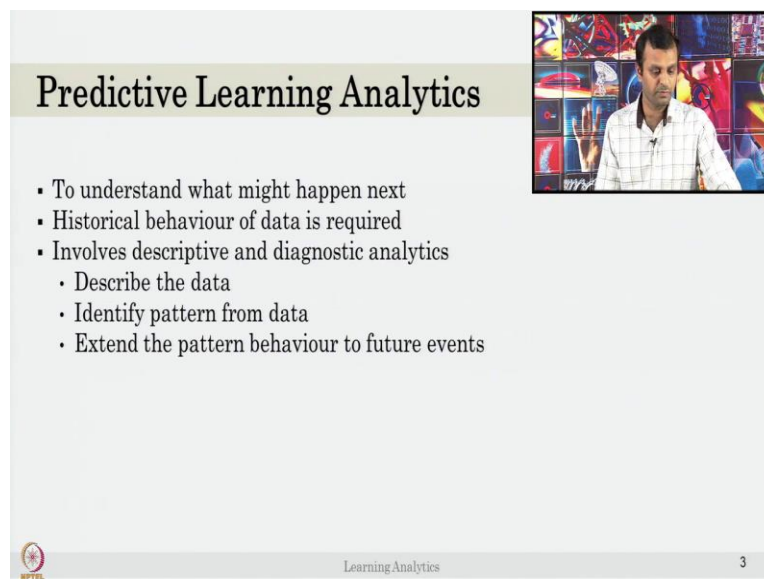
(Refer Slide Time: 00:29)

Welcome back to Learning Analytics Tools course. In this week, we will discuss Predictive Analytics. We saw in first week that there are four levels of analytics in learning analytics and we discussed very briefly about descriptive analytics. It is about pictorial representation, and how to represent the data. For last three weeks, we had seen what is diagnostic analytics, we discussed pattern mining, correlation and some clustering techniques. So, i hope you understood how to perform diagnostic analytics. The next step is predictive analytics and that is the last step of this course. We do not talk about prescriptive analytics in this course.

So, when we say predictive analytics which means it will measure what will happen next. Diagnostic analytics is about what happened and why it happened. Predictive analytics is about what will happen, and if you know diagnostic analytics you can create a relationship between dependent and independent variable, you can use that relationship to create the future events that is called predictive analytics. So, you can use that relation to predict the future events.

So, predictive analytics subsumes diagnostic analytics and descriptive analytics. So the first step is doing descriptive analytic, i.e. looking at the figure, checking the out layers, figuring out relationship from the plot if it is possible. Suppose, if we have more than say 10 variables it is not easy to plot, you can get the sense of each variable data and the relationship, but you would not get a complete picture then you can use the correlation matrix, to know that is there any relationship. Then you can go for the predict analytics to create a model which can predict what will happen next.

(Refer Slide Time: 01:58)

So, predictive analytics is to understand what might happen next. And this is based on the historical behaviour of data. So, in order to predict what will happen next, we need a historical data to train the system as well as to test the system. It involves a descriptive and diagnostic analytics that we shown in the previous slide and which means you have to describe the data, identify the pattern from the data or identity the correlation relations from the data, then extend the pattern, to the future events. So, that is the idea of predictive analytics.

(Refer Slide Time: 02:34)
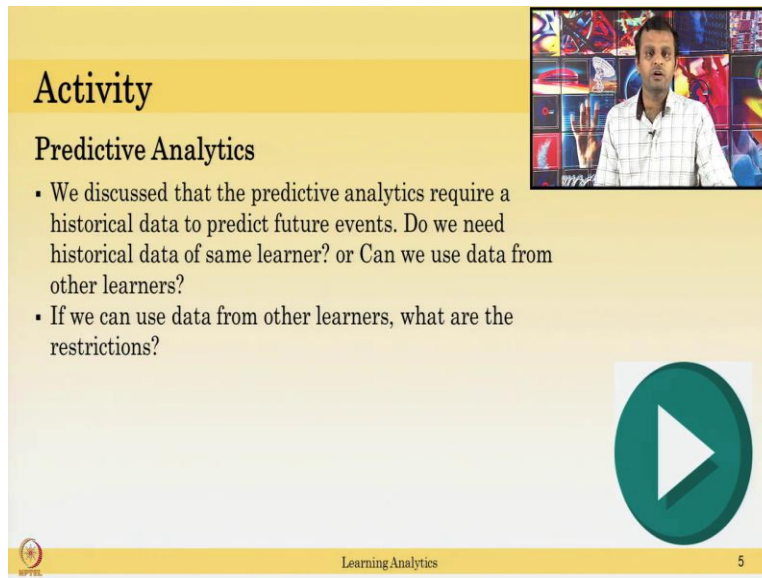

So, predictive analytics analyse the current and historical data to predict the future events. Suppose, if you are teaching in a class this year, then you might use the data from this year students and also data of the previous year in order to predict the final exams. Since you do not have the exam score of this year students, as they are currently studying. Also we will use machine learning and data mining tools to create a predictive analytic models.

For example, other than in education domain, this has been used widely, that is in weather predictions, fraud detection in credit card or loans, anti-spammers in email, in domains of finance it can predict the credit card risk or loan risk, in health sector in order to understand the insurance predicting whether insurance will get paid or not so that we can ask the more premium, and in telecom sector it is used to predict which user will drop out, or which user can go for the higher package.

(Refer Slide Time: 03:54)

So, we have discussed that predictive analytics require a historical data to predict the future events. Let us, talk about a educational domain, in educational context, do we need historical data for same learner? Or can use data from other learners? Can you think of that? Also if you can use the data from other learners, what are the restrictions? So, we said that we need a historical data to predict future events, so what is a historical data? Is historical data for same user or other users? Think about it, after you write down your answers, resume to continue.

(Refer Slide Time: 04:32)

Activity

- We may not be able to collect enough data from a learner to predict their future events. Hence we need to consider the historical data similar learners.

- Similarity
  - Domain
  - Interaction behaviour
  - Interaction time
  - Data collection

We may not able to collect enough data from the same learner to predict the future events. You know the learner has joined newly in a first year of the college or the learners are first time interacting with intelligent tutoring system, as you might have created a system to teach one specific topic.

But you can use the similar learners from the previous study or previous years of data. So what should the previous learners have to be similar, and what is the similarity basically? For similarity they should be from same domain. For example if you are collecting a student's data in class say mathematics of second year students, the you should have the last two years or last three years students who attend a mathematics in second year.

Or also they should have similar environment. The interaction behaviour, either in classroom environment or the kind of teaching strategy used should be similar. Also, if you are introducing a new teaching strategy that can vary, except it all other things should be same like, a classroom and students pre and post test score or their ability everything is same. Their similarity should make sure by doing some statistical measures.

Also for the similarity, the interaction time should be same, it is not that in one year student using the system only for 30 minutes and next year you are collecting data where the student is interacting more than say one hour. So, make sure you have a same domain similar type of

interactions, similar interaction time with the system. So that that kind can of data can be used to predict the future events.

And also, the data collection method should be similar. It is not that you are collecting extra data from the previous year to this year. But if your research question is valid and you want to collect new data to prove something extra it is allowed. So, what I mean here is if you want to use the historical data, please make sure that you have a similar domain, similar setup of study and a similar data collection method.

For example, in intelligent tutoring systems, we might created a system to teach a concept say Newton's law. We have some method to teach and there also I have a set of assessment questions to assess the students. I created a system to teach Newton's law to say class eight students.

Now, you collect data from say 100 students in one school where students are interacting with the system .You collected the data, you are not doing any prediction because that is a new system, or it is a pilot study. So you created a system and you are going to establish whether this shows a learning gain compared to pre-test and post-test when students learned due to your system, your intervention, compared to normal intervention, like classroom education. Then you establish that the system is good.
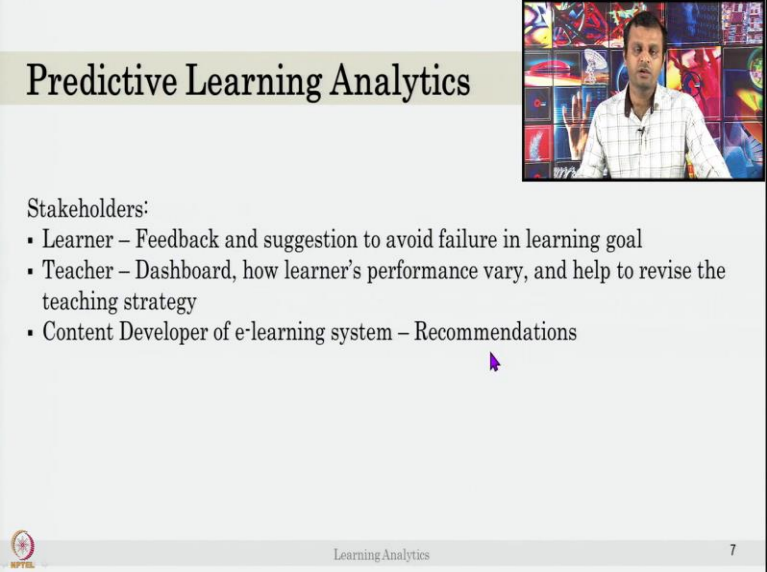
Now, you want to create a predictive models to predict which user will struggle. Can you provide more hint to the user something like that. In order to do that use this 100 students data and create a model to predict the interaction with the user.What will be their mark in assessment questions, whether a user can understand the concept or not.

So, when we go for the next set of study, we have to make sure the users have similar pre-test level, we can run a statistical significance test to know whether these two group of users have the similar prior knowledge. Then you check the interaction time if one group of user taken one hour, then other group should also use one hour. Also you do not introduce any new element here, you make sure then you conduct a study to test whether your system works or not.

By running multiple studies like that you can say that your system can predict the student's interaction and that can be useful for the providing recommendations. It is not that 30 students in

the first study, itself we can provide hint and recommendation. That maybe from the hypothesis you are giving and not for the data.

(Refer Slide Time: 08:46)

So, predictive learning analytics vary based on the stakeholders. For example learners, if this will be helpful for feedback and suggestions to avoid failure in the learning goal. For teacher it will have dashboard, and it can monitor, how learner performance vary and can also help teacher that which learner having struggle, or so whether the lot of students in the class are not able to understand particular concept so teacher can use that data to the revise the teaching strategy. Also in a content developers of e-learning systems, they can use this predictive analytics to provide recommendations, hints or adaptive content on that.

(Refer Slide Time: 09:27)

In this video we discussed, what is predictive analytics and what is the importance of historical
d
a
t
a
.

T
h
a
n
k

y
o