Learning Analytics Tools

Professor Ramkumar Rajendran

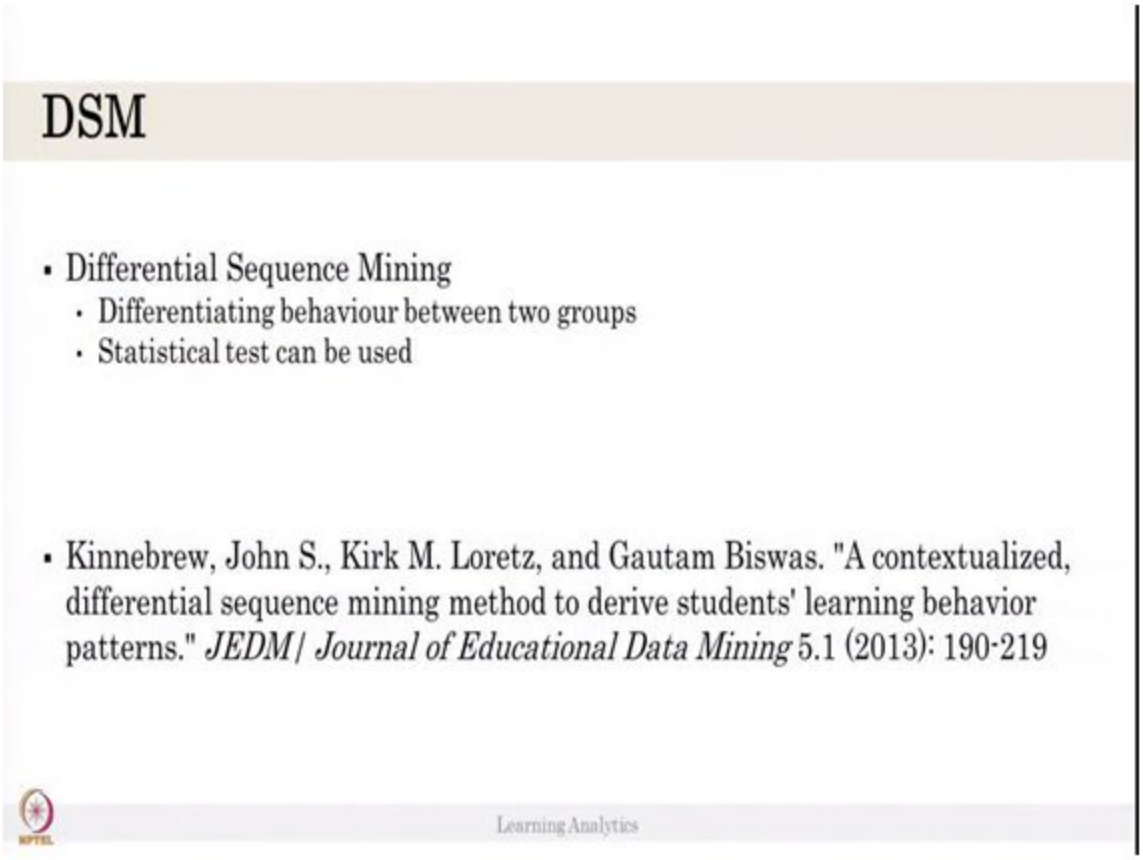
Educational Technology

Indian Institute of Technology Bombay

Lecture 34

DSM

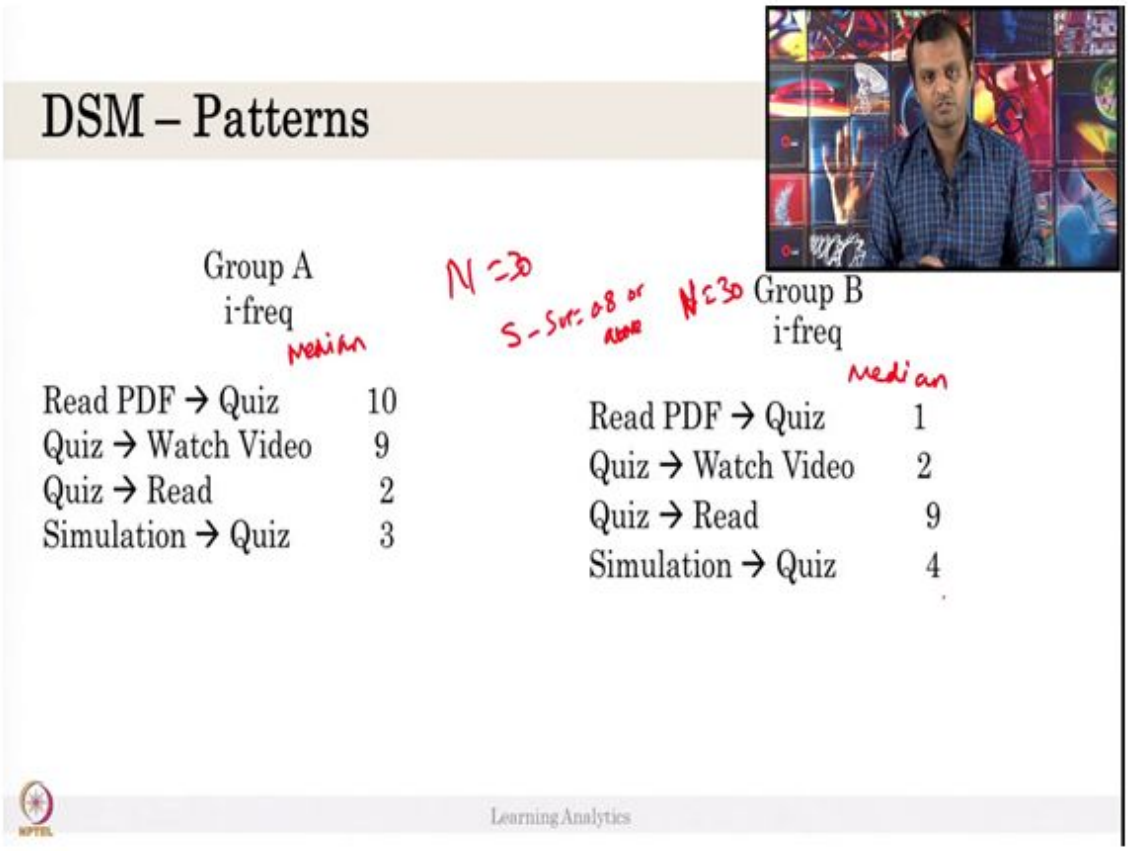
(Refer Slide Time: 0:28)



In this video we will discuss differential sequence mining, this is an extension of SPM. So, differential sequence mining is trying to identify the differentiating the behaviour between two groups by using the patterns, like the patterns identified in the SPM algorithm. So, we can actually use a statistical test to identify whether this is statistically significant or not. That will

help you to make a better claim. And if you want to know more about of this differential sequence mining, how it is used, how it started, study this paper 2013 paper, from a journal of educational data mining.

(Refer Slide Time: 1:01)



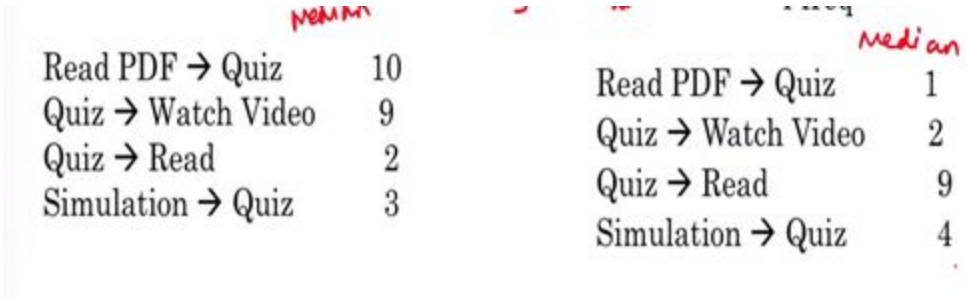
Let us see, what is DSM? I will try to explain it with a very simple example. So, there are group A group B, consider Group A is from one section and other section is going to Group B, you are teaching the same course, all the students are equal and consider all assumptions carried out.

And you did all the tests to make sure everybody is equivalent or within the same class, you can split into two groups and you are giving this kind of intervention. You are asking them to interact with the MOOC, or you are creating a simple learning environment where they have to watch the video, read some GUI pages to answer some questions.

They are free to do everything and they are free to navigate in the particular environment, you ask students to solve a particular problem, then you captured all the log actions like the reading action gui's actions. Then you want to identify the difference in behaviour of the group A versus group B. It is very simple to identify the behavioural differences but if you are doing pattern mining, let us see the group A did "read to quiz" 10 times consider this is i-frequency median.

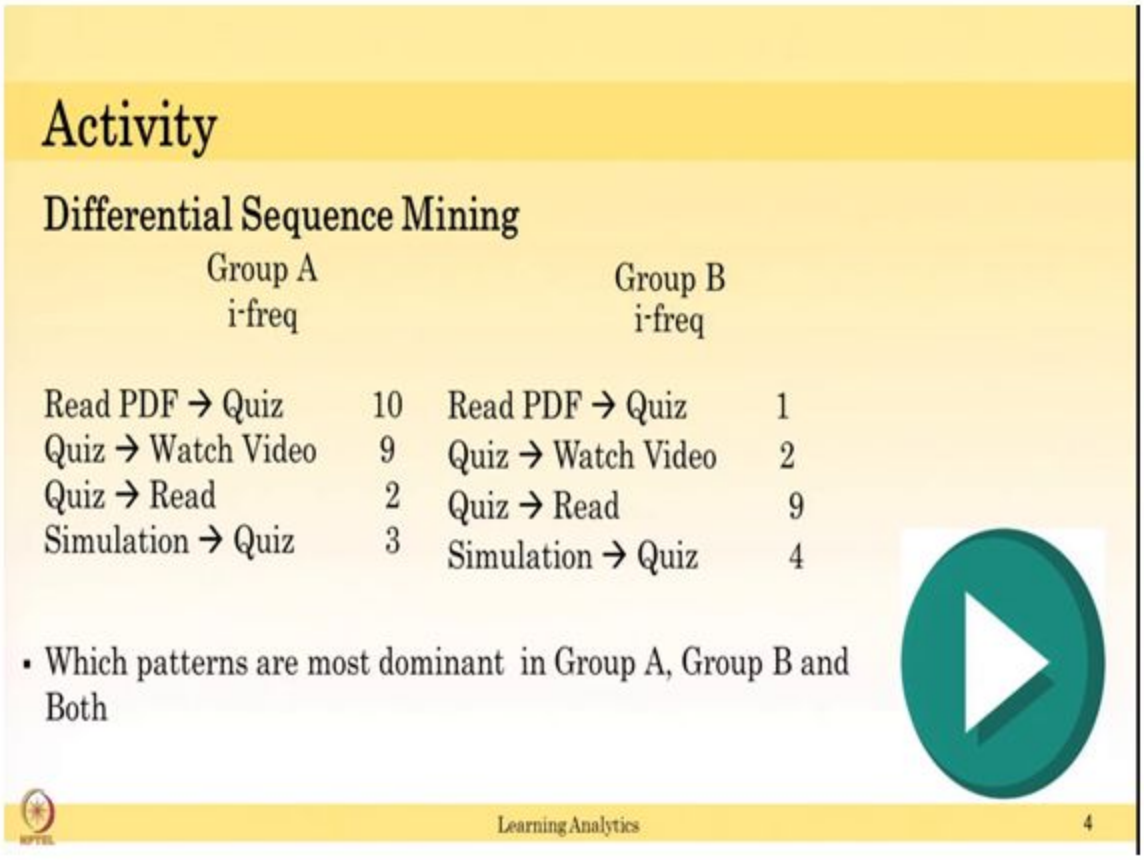
And imagine the s-support is equal and say 0.8 for both, 0.8 or 0.6, N is not important here consider the N is also equal to say 30 or something like that and s is 0.8. Consider the N is 30 in both and s support is 0.8 or above, so, you have i-frequency, it can be a median, let us say its median. Not a mean.

Group A	Group B



So, you have this median values i-frequency values. Can you identify the difference in behaviour between these two groups? It is obvious that this group(A) had more "read to quiz". So, can you identify other differences in these groups?

(Refer Slide Time: 3:44)

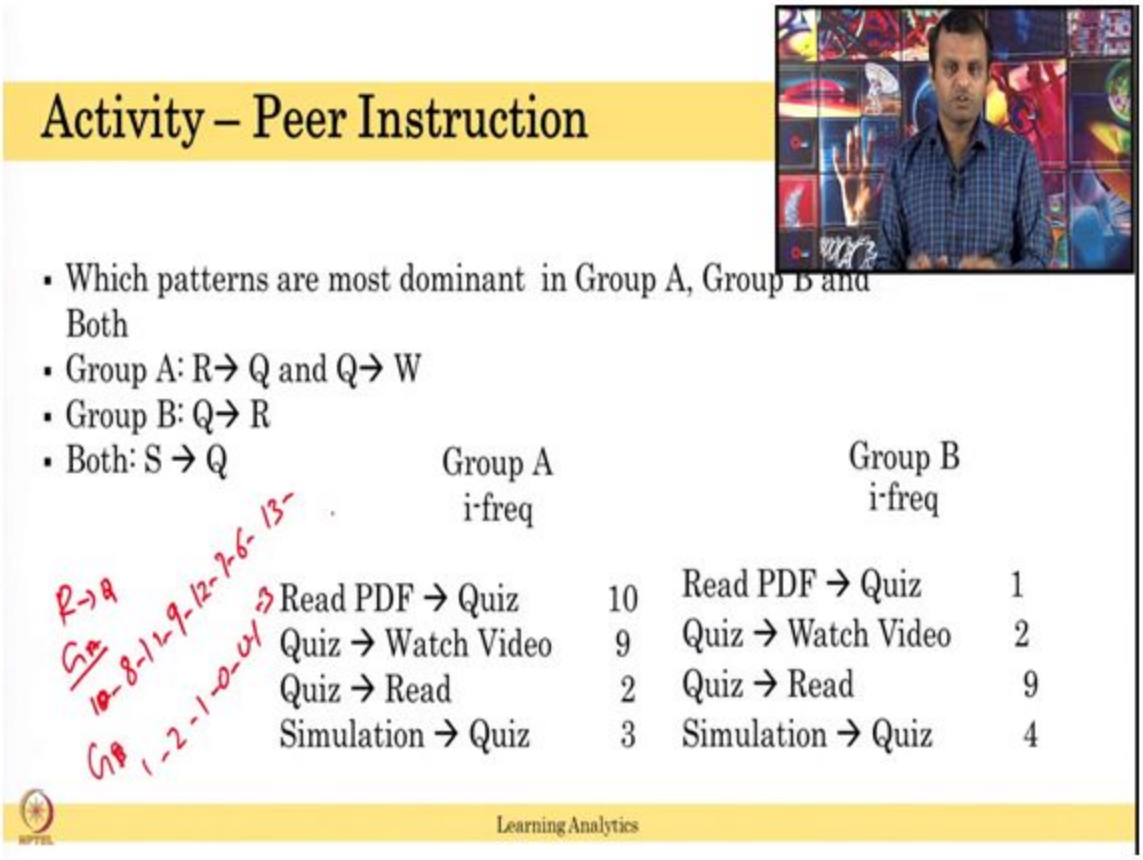


So that is an activity. It is easy to identify, I just gave the answers. You can talk about which patterns occurred more frequently in Group A which pattern occurred more frequently in Group B, which pattern kind of occur, similar in both groups this is what I want you to think. So, there are four patterns classify into three groups.

A. The pattern which occurred more in Group A
B. The pattern which occurred more in Group B
C. a pattern which occurred in both groups.

After, you list down each individual to continue.

(Refer Slide Time: 4:21)



So, it is simple Group A had a pattern, “read to quiz” and “quiz to watch video”. This is simple you can compute. And Group B it is “read to quiz” occurred more compared to group A, “read to quiz” and “simulation to quiz” occurred kind of both groups. This is a basic of difficult differential sequence mining. It is very simple if you have only two groups one of these patterns is easy to identify.

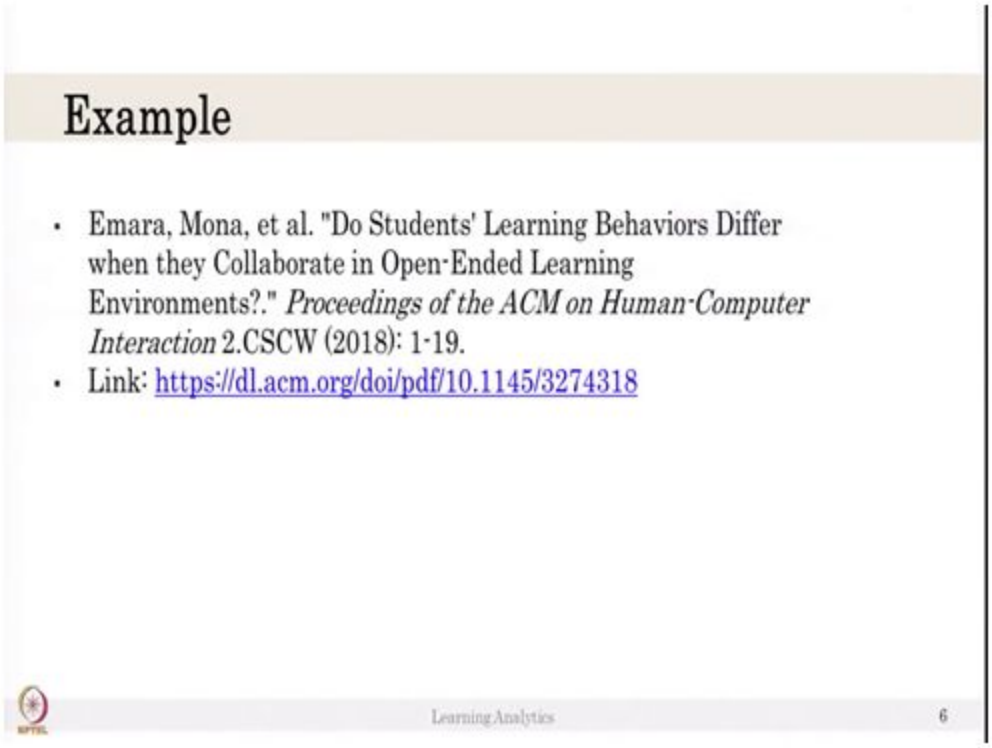
But if say if I am to use the statistical significance test on this particular patterns and differentiating this I want to see is there any statistically significant difference between these patterns occurring more in Group A compared to Group B. If you want to do that, what you have to do? You remember the i-frequency, we computed use this value to compare the statistical significance.

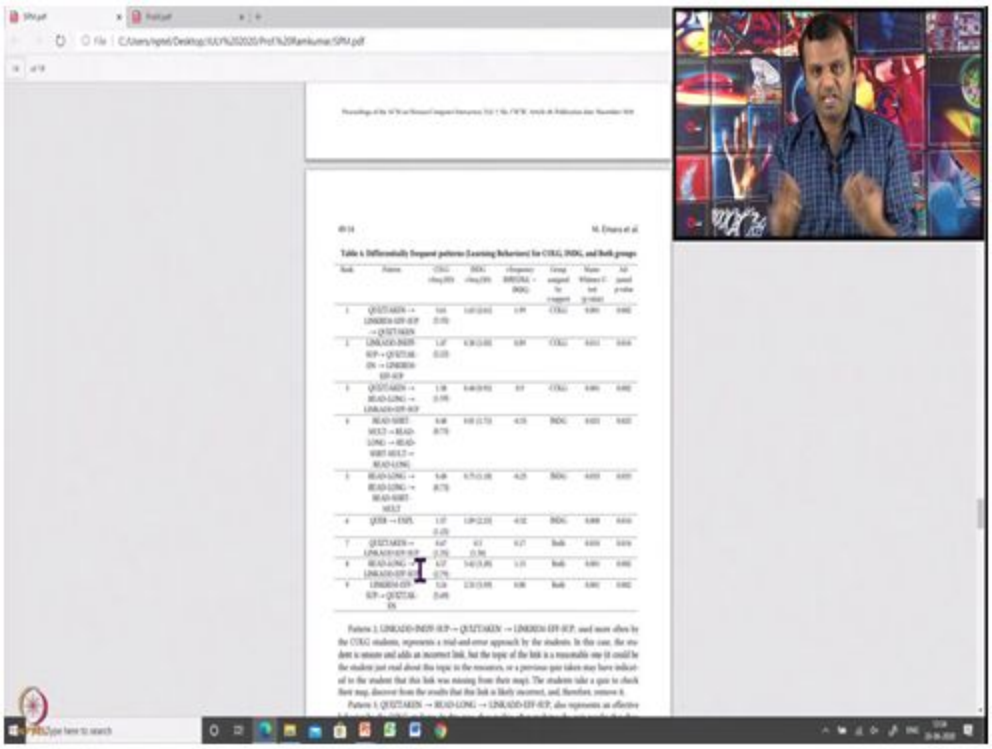
Let us say this for “read to quiz”, suppose this value read to quiz for group A, it occurred 1, not 1-10-8-11-9-12-7-6-13, it occurred in this kind of sequence. And for Group B, it occurred in

something like this. There are two sequences you can use these two sequences to create a statistical significance test, like a T-test, or Mann Whitney test.

So, you have to be careful about which test you want to use, so it is not normal. It would not be a normally distributed, so maybe Mann Whitney test, you think about that, that is not part of this course. But yeah, just go and read it, just for fun understand which test to use on which kind of data on statistical significance. So, yeah, so if we have a statistical sequence used on that, then that may tell a better value. So, that is the basics of differential sequence mining.

(Refer Slide Time: 6:36)





So, let us see the same paper which we discussed. We saw that in our last class in a bit more detail. The paper we saw in the last class, the same set of actions. I want to go back to the same

table. This is what happened here. So, what happened? There are around 40 50 students who worked on individually, there are 24 groups who worked on the collaborative group, you can read the paper to understand the exact number and there is a difference between the frequency suppose collaborative group have this more, compared to the individual group.

So, they computed the Mann Whitney U test to find or identify whether these occurrences significant or not. So, this particular occurrence is significant, this particular pattern occurred more in the collaborative group compared to individual group similarly for these three patterns occurred more in collaborative.

And these two, three patterns occurred more in the individual group compared to this collaborative group and these patterns occurred in both. With these values, you can make a new inference about the behaviour difference between these two groups. For example, you can say the collaborative group students mostly had the support you link and take a quiz and they remove that ineffective links compared to individual group students. They simply read read read read read, which never occurred in the collaborative which occur in very less frequency.

So, the individual group who do not have someone to talk they are simply reading, reading, reading and trying to understand the meaning on their own. Whereas collaborative group students will not read much because they can collaboratively work together, 2 students are working in a collaborative group, 2 or more so they are talking and the understanding they are able to solve the sums easily.

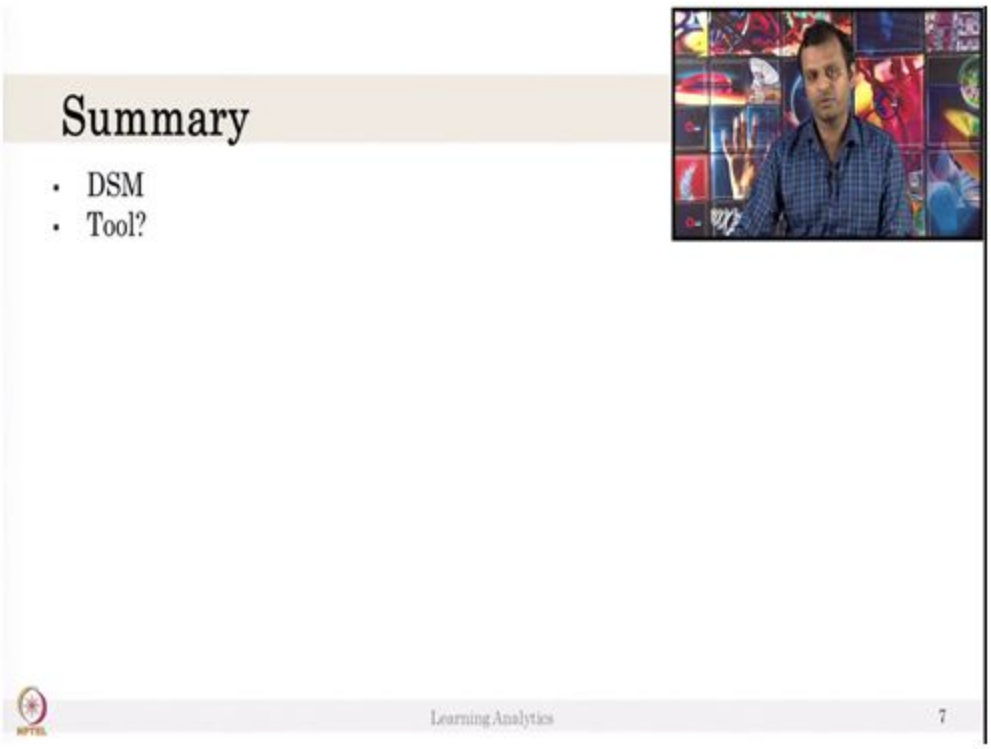
So, the individual group might be always asking query and explanation, you may be asking the agent what is this, ask the agent to talk because they do not have anyone to talk. So, they will be asking agent a lot of questions that particular pattern is more dominant throughout in individual group compared to the collaborative group. So, these kinds of behaviours will tell you, how to create inference from these particular patterns.

And there are some of the patterns which occurred for both that also tells you what is the common behavior between these two groups, if you are conducting studies, say in a class of 60 students, you want them to work on the online environment and you can do a pre-test and post-test and based on the scores, if you want to classify them as high score versus low score,

you can create as a two-group - high scoring students, low scoring students, then you can run a pattern mining based on the actions to see, is there any behaviour differences.

In diagnostic analytics, we try to understand, why a student got less score compared to the other students who got a high score. Here we are not trying to predict anything. Instead, there is one group who did really good, what is their behaviour, what is interaction behaviour in the system, compared to the students in the low group of what is the behaviour the low Group did very good, but high group students are not able to do, this kind of analysis, diagnostic analysis can be done using simply the pattern mining or differential sequence mining.

(Refer Slide Time: 9:56)



Hope you understood what is SPM and DSM in the last two videos, and we do not give a tool for DSM instead we give the frequencies, values, i-frequency, so use those values and compute your own DSM behaviour and compute your own a statistical significance test. Since the data is not normal, I will suggest Mann Whitney use, best to use.

But, there are a lot of tools available in online to compute Mann Whitney U. So, go ahead and check the tool this week and compute some patterns from using your own data or create your own data and test it out and read about which test significant test to use like Mann Whitney U, why and apply it and understand what is SPM and DSM. Thank you