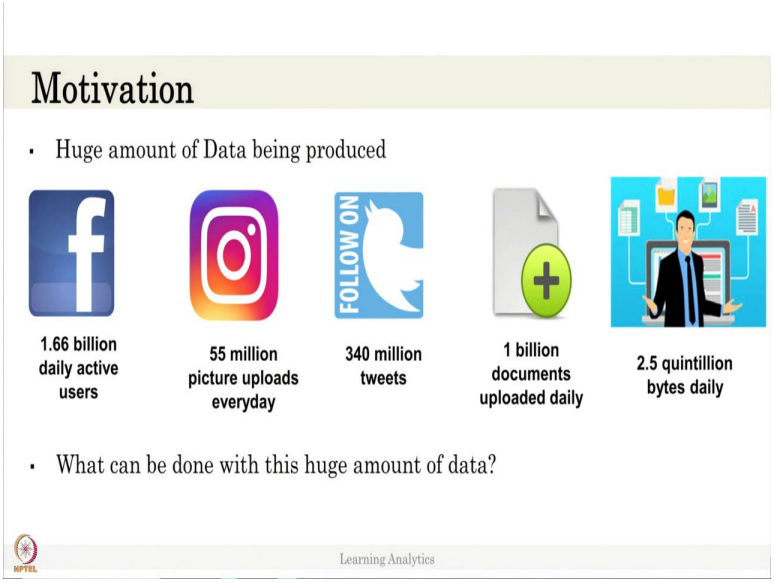
Learning Analytics Tools
Professor Ramkumar Rajendran
Department of Educational Technology
Indian Institute of Technology, Bombay
Lecture – 1.1
Intro to Data Analytics. What is learning Analytics?

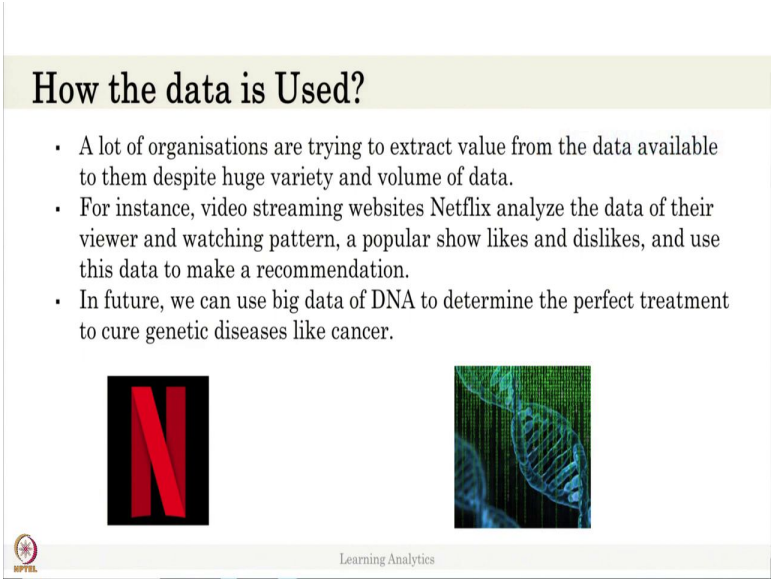
Hello and welcome to the Learning Analytics Tools course in NPTEL. This will be a two weeks course. I am Ramkumar Rajendran, assistant professor at Educational Technology department of IIT, Bombay. I offered the previous course called Introduction to Learning Analytics, a four week course prior to this course. If you have done that course, you can see the same content or similar content for first two weeks; however, this course does not require any prerequisite. So, even if you are not done the introduction to learning analytics course; this course you can start. Due to web two point O, the users can generate lot of data.

(Refer Slide Time: 1:08)



For example, in social media networks like Facebook, there are 1.56 billion monthly active users, also at Instagram and Tweeter a lot of data has been generated. But, what can we do with this data?And more interesting is to know. How to use this data?

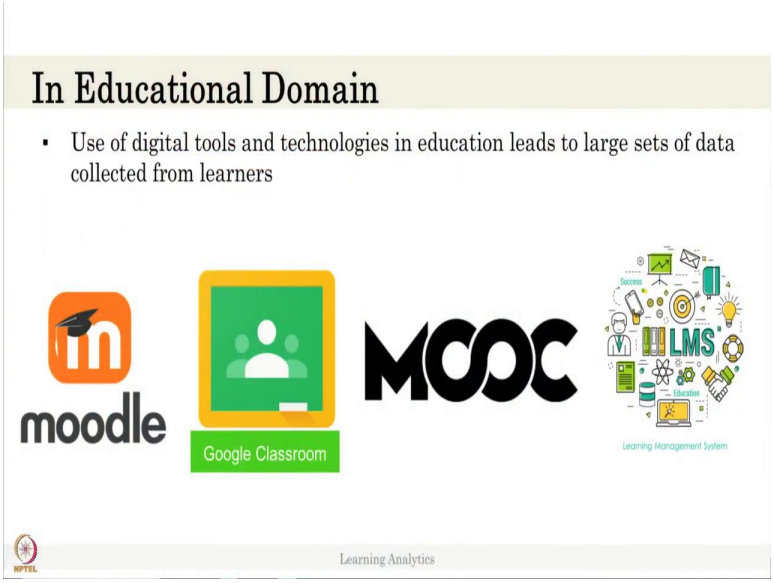
(Refer Slide Time: 1:14)



For example, a lot of organizations are trying to extract values or patterns from these data. And cloud computing technique and a number of servers are used to understand what is user behavior. For example, in Netflix the video streaming websites, the data for viewer's users viewing behavior, which movies they watch, their ratings, the likes, the dislikes are used to recommend the next video further.

Similarly, in e-commerce websites like Amazon, based on purchases, the system can recommend what will be the next product to purchase. In future we can use big data of DNA and the systems can come up with the cure for disease like, Cancers or Malaria. It is possible because now the lot of data available of the humans DNA and also a lot of medical data is available. We can use those data and we might get the cure for these diseases.

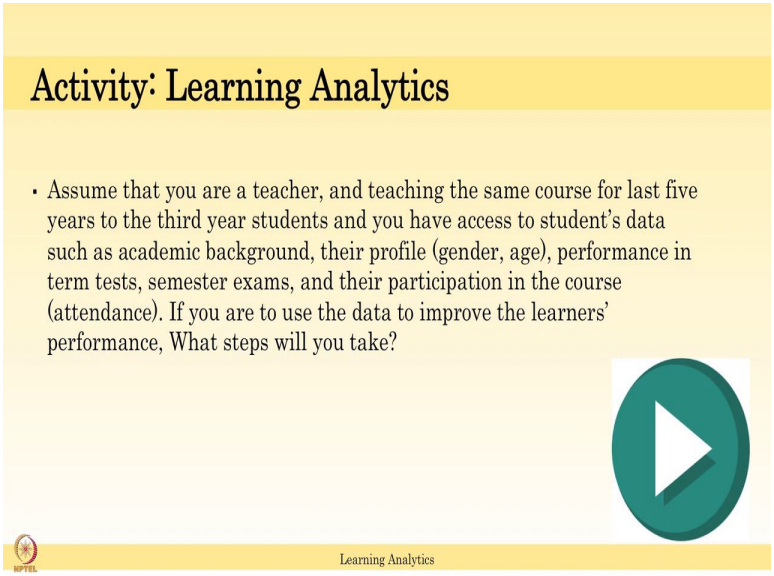
(Refer Slide Time: 2:20)



Like in other domains in education domain also, a lot of data has been generated . It is because of use of digital tools like learning management systems, like Moodle; a number of educational apps, Google classrooms, massive online open courses like MOOC's in course like NPTEL or MIT course, where a lot of this data has been generated. The user's interaction with these interfaces has been collected and stored.

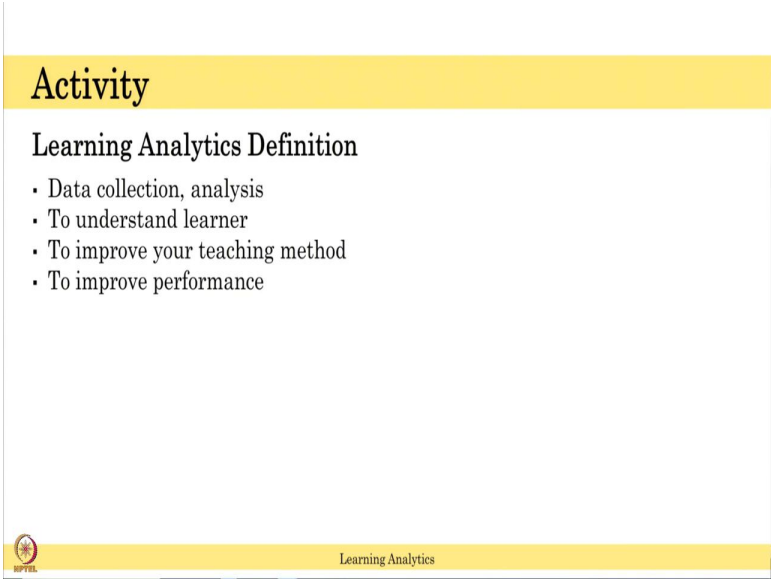
What can we do with this data? Let us start with the first activity.

(Refer Slide Time: 2:53)



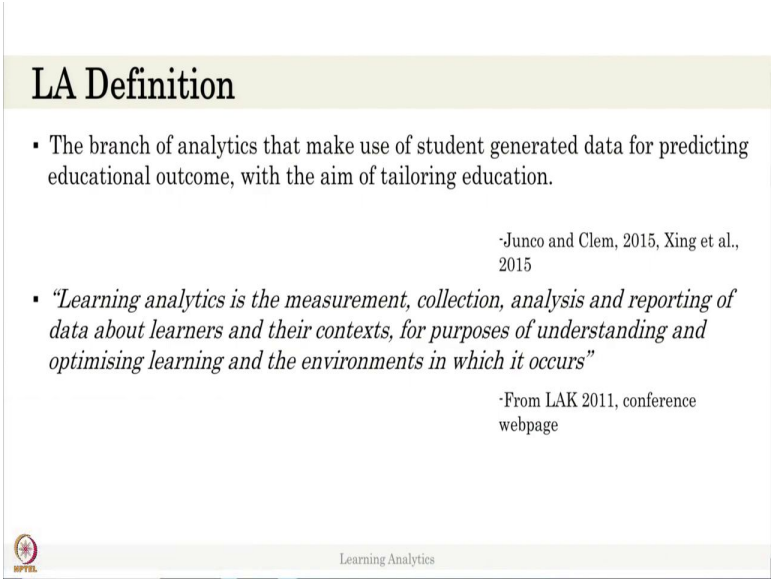
Assume that you are a teacher; or you may not be a teacher, but please assume that you are a teacher. Also you teaching the same course for the same class for last five years to the third year students. And, you are teaching third year electronic devices course for last five years. So, you have data of the students for last five years, their background, their profile, their performance in the exams, mid-terms, performance in assignments, feedback in the classroom. So all this data you might have it. So, if you have to use this data to improve your teaching strategy or to improve the learners' performance, what step will you take? Please pause this video and think about your answer. Write down your answer, after writing your answer and resume the video to continue.

(Refer Slide Time: 3:51)



If your answers contain the words like, I will collect data, I will analyze data or I want to understand the learners learning or I want to improve my teaching performance so that the learners can learn better. If you have these kinds of words or sentences, then you are thinking in learning analytics. You have already started thinking, how to analyze the data and can improve the students learning performance.

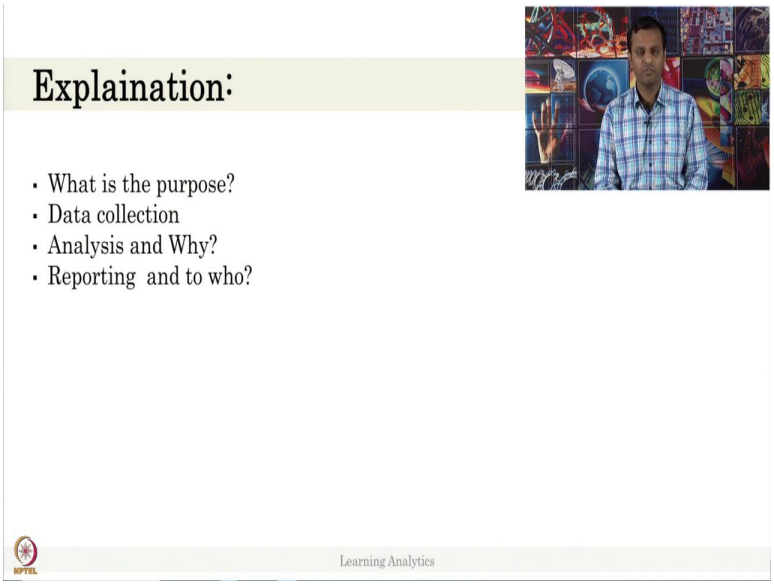
(Refer Slide Time: 4:22)



Let us see what is a formal definition of learning analytics? The learning analytics definition is still a debatable one. But, from the existing resources, it can defined as "Branch of analysis that makes use of students generated data for predicting educational outcome, with the aim of tailoring education with the aim of adapting the content, so that student can learn better".

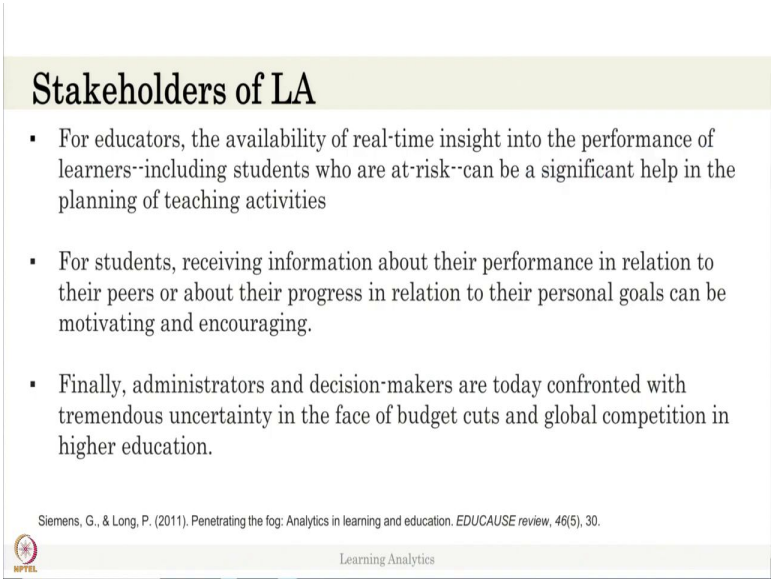
In LAK 2011 conference, which is started by organization called SOLAR, the learning analytics definition is as following "Learning analytics is a measurement, collection, analysis and reporting of data about learners and their contexts, for purpose of understanding and optimizing learning and the environments in which it occurs." What does it mean? It says learning analytics is measuring, collecting and analyzing the data; not just analyzing for the analysis purpose, but also for reporting this data to a stakeholders of the LA, reporting about learners and their contexts. The sole purpose is to understand how learners learned in that environment and what can we do to improve the learning in that environment.

(Refer Slide Time: 5:42)



In this definition there are few terms. What is the purpose? The core purpose is to understand the learners learning process and help them so that they can learn better in the environment. Data collection, what data to collect? How to collect data in that environment. How to use those data and analyze? What to look for in that data? And why we have to analyze? Also if we want to report the data to whom we have to report the data? So, who are the stakeholders of LA?

(Refer Slide Time: 6:14)



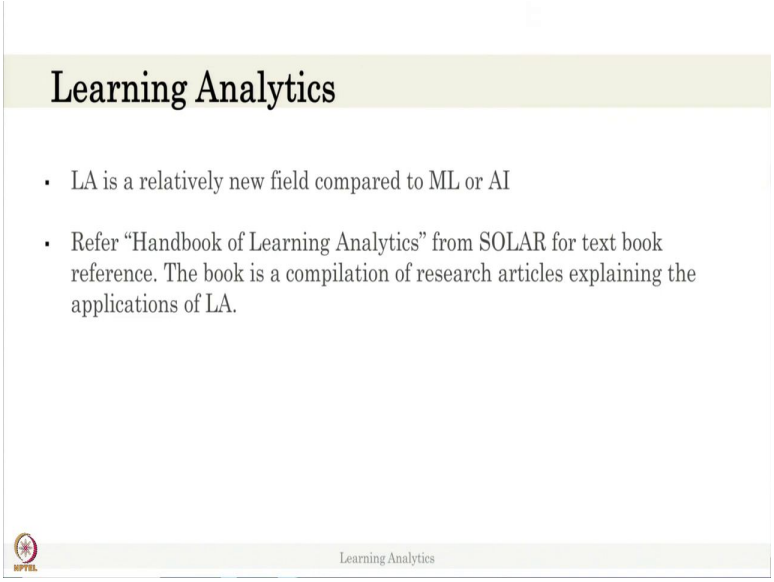
For example, the stakeholders can be educators, that is the teacher or the instructor. They can use the availability of real-time insight into the performance of learners. For example, if the teacher has a dashboard of the students working on a particular exam or a particular learning environment. All the known interaction is given to the teachers by a dashboard. A teacher can have a real-time insight into the performance of learners including students who are at-risk, and the teacher might know this particular student needs help. The teacher can go and help them or the teacher finds out that whole class is having misconception in a particular topic. So the teacher can teach the topic in a better manner or give a remedial content to that.

Similarly, students for example, receive information about their performance compared to their peers in the classrooms; or their progress compared to the peers in their classroom, that can help them to motivate and achieve their goal. For example, if a student answers a question wrongly; and he thinks that the option might be correct. But, if it is shown to the student that in the class, almost 40 percent of students selected this particular option and they are wrong. The student might feel that I am not the only one who has given a wrong answer. So, everybody did not understand this concept; so I can learn and the students get motivated to continue. If the student thinks I am the only one who do not understand anything in the class, then they get de-motivated. They might or might not in getting interested to continue further. Also the students want to know, how much I progress in this particular course. If the student knows that I progress 40 percent of content in this course and I need to cover another say 20, 30 percent. Student can think that, I need to put a more effort to continue this course or the student can decide whether to continue the course or drop the course. So, the learning analytics can help the students motivate them, and encourage them to complete the course and also help them to compare their performance with the peers.

Finally, the learning analytics data can be helpful for the administrator to make decisions. In the today competitive world and reduced budgets. The administrator should take a decision that should provide optimal solution that is, whether to run a course or not. If the administrator knows that particular course receives very less number of students, and they know the reason and if they can predict how many students will join the course in the next year, then they can make a decision whether to run the course or not. They should convert the course into some other

topic, because it will save a lot of cost in today's world. Hence during the global competition in higher education; it is always useful to look at the data and make a wise decision.

(Refer Slide Time: 9:48)

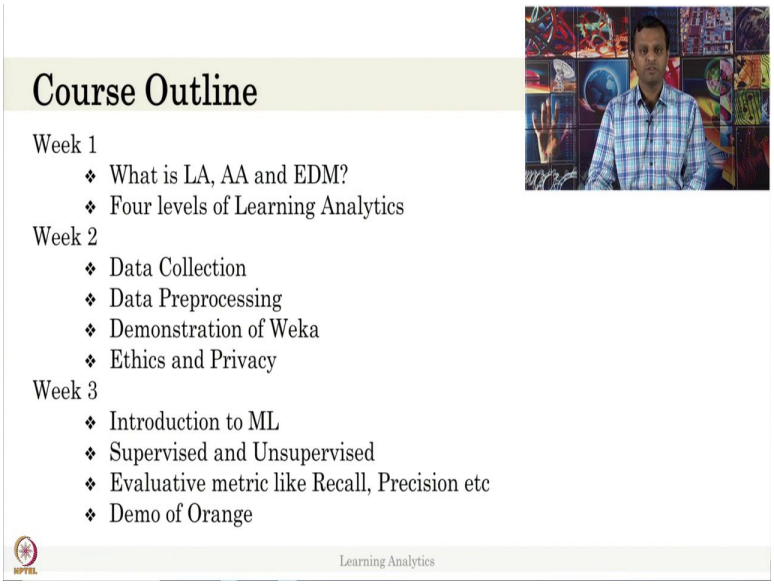


As you know that learning analytics is relatively a new field even compared to the other new fields like a machine learning or artificial intelligence. Since, it is very new field; there is no standard textbook available for learning analytics. However, in this course we will cover the basics of learning analytics in terms of analytics applied to education data.

Then we will explain some tools useful for this course to analyze the data. We will also refer to some content from this textbook, "Handbook of Learning Analytics" from the SOLAR community. The book is actually compilation of research articles, explaining applications of LA in different fields. So, when you read this book, you can understand that every chapter introduces something about learning analytics.

So, if you are interested in learning analytics after this course; please read this book or read our papers in the recent conferences. So that you will understand the fields in learning analytics, then you can pick one field which you like to pursue your research interest in LA.

(Refer Slide Time: 11:10)



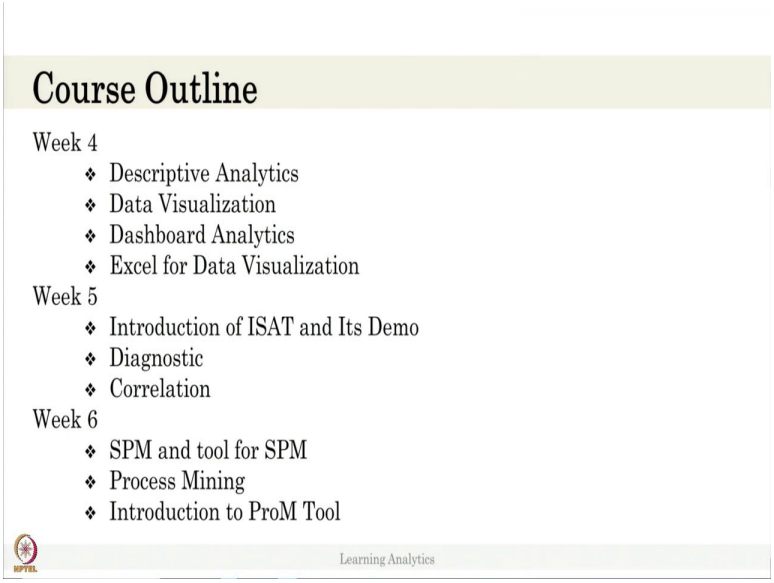
Now, I will briefly describe the course outline in this video. In week , I will be discussing, what is learning analytics, or academic analytics and what is relation between it and with educational data mining? Then I will introduce what are the levels in learning analytics? There are four levels in learning analytics and I will explain with examples.

In week 2 I will talk about the data collection, what data to collect in each environment. For data pre-processing, we might give you links to the external sources; so we want you to go and study what is data pre-processing from the others resources. And we might have a assimilation quiz based on that reading exercises. Also in week 2, we will introduce a tool called Weka as a freely available, open source tool.I will demonstrate the tool Weka, which we will use it in our course going further. I also will introduce the ethics and data privacy that is when you collect data from the students, what are the ethics you should follow.

In week 3, we will introduce the basics of machine learning. This course is not mean to teach machine learning; also the course may not involve mathematical details of each algorithm. So, this course is designed for anyone with the very little mathematic background to understand how

to collect data, how to start doing analysis. So, in week 3, we will introduce what is machine learning, what is supervised and unsupervised machine learning and very basics introduction to what are the metrics you should look for when you use the machine learning for your data. Also in week 3, we will introduce a tool called orange. This tool is not free for commercial purpose, however, academic users, that is if you have academic email id; it is free to use. So, we will demonstrate the tool called orange. If you do not have access to that tool; you can continue using Weka in this course.

(Refer Slide Time: 13:14)



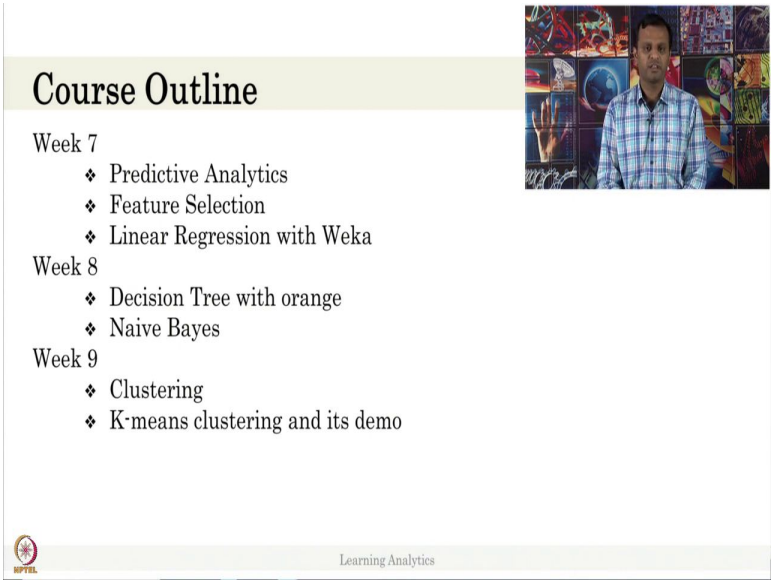
In week 4, we will introduce descriptive analytics that is how to describe the data, and what is data visualization and how to look at the data. How the data is generated in dashboard. We can use Excel or Google to produce all these visualizations from the data.

And in a week 5, we will introduce another tool called iSAT. This tool is used for visualization, also for diagnosis purpose. This tool is developed by our departments, so will demo that tool; which will help you to visualize the data transfer from one stage to other stage in different time periods. And we will also talk about the diagnostic analytics and diagnostic analytics starts with correlation, regression. .

In week 6, we will have a sequential pattern mining and the tool for sequential pattern mining will be described. The tool will be available free and we will upload the links to this tool, and

anyone can use this tool; if the data is formatted in right format. We will also provide the demonstration of the tool. Also, we will introduce another tool called process mining; the tool called ProM. This is also freely available for educational purpose; so anyone can use it, if you have educational id. Also, the ProM is available for everyone; I think its open source.So, by week 6 will be introducing several tools like Weka, Orange, iSAT, SPM tool and ProM. So, five tools we plan to introduce in this course.

(Refer Slide Time: 15:04)

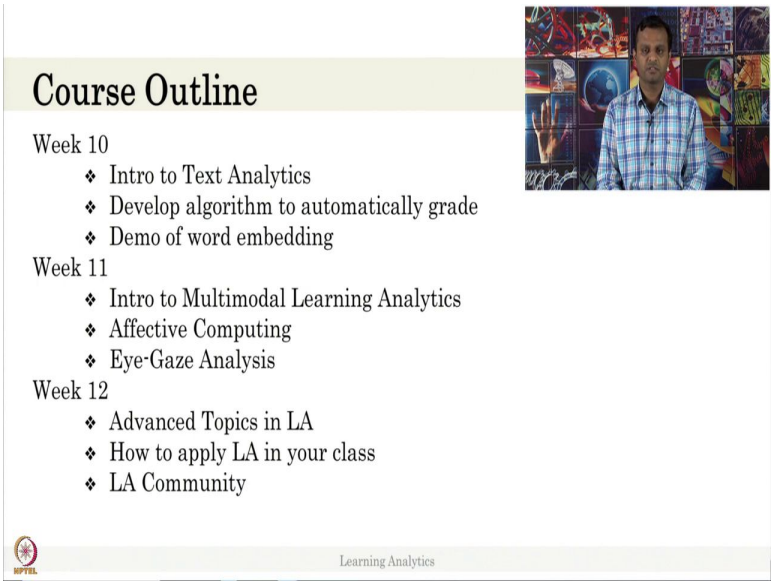


So, after the mid-semester break, in week 7 we will talk about predictive analytics and we talk briefly about what are the features to select for linear regression using Weka.

In in week 8, we will talk about decision tree and we will explain this with the Orange. So, whenever we have a demonstration, when we are talking about explaining a particular algorithm with a tool; which means we will have a course assignment. So, you have to use the tool and we provide the data; using that data you have to predict something, or you have to create something and report that as answers. So, when we are talking about this demonstration of tools will be having assignments in each week. And in week 8, we will have described what is decision tree with orange and we will talk about a Naive Bayes. As I mentioned this course although it is not ML course; we will touch all the algorithms which are very basics, we will explain the concepts of how it works. Then we will also show how to use tools to execute these algorithms.

In week 9 will go for the unsupervised machine learning that is clustering and we will show, what is clustering and with the demo.

(Refer Slide Time: 16:28)

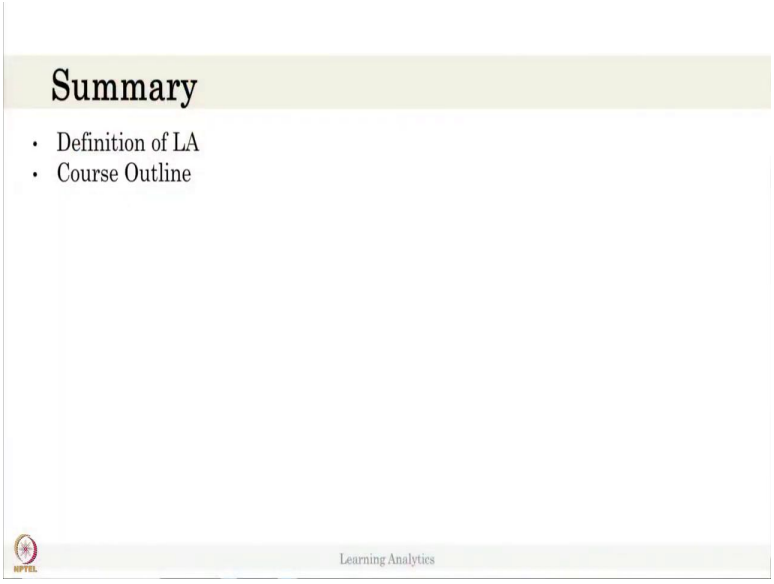


In week 10, we will jump to a different field called text analytics or natural language processing applied for educational domain. So, we will show you how we can use text analytics to develop algorithm that can automatically grade student's essays.Also we will talk about the latest development in NLP that is called word embedding, or word vectors.

And in week 11, we will talk about multimodal learning analytics; that is how to collect data from multiple sensors like Eye-Gaze data from eye trarkers or facial expressions for using webcam, log data or biosensor either like EEG or EMG or GSR. How can you collect these data, use these data and analyze this data to create a model. So we will show a bit about how to collect data; what is this data is used for and what is this model looks like.

In the final week, we will teach about advanced topics in LA. It is basically covered in the handbook of LA, we also include also latest topics published in LAK conference and EDM conferences.

(Refer Slide Time: 17:54)



So, that is the course introduction; so in this course we briefly described, what is definition of LA or LA means? You have to collect data and analyze data. Then you have to report data to stakeholders to improve the performance of the students. Then we described the course outline; the course outline involves five tools; a lot of assignments and a lot of data to collect and a lot of exercises in it. And that is all for the motivation video, this is a first video. Thank you for watching.