Learning Analytics Tools

Professor. Ramkumar Rajendran

Department of Educational Technology

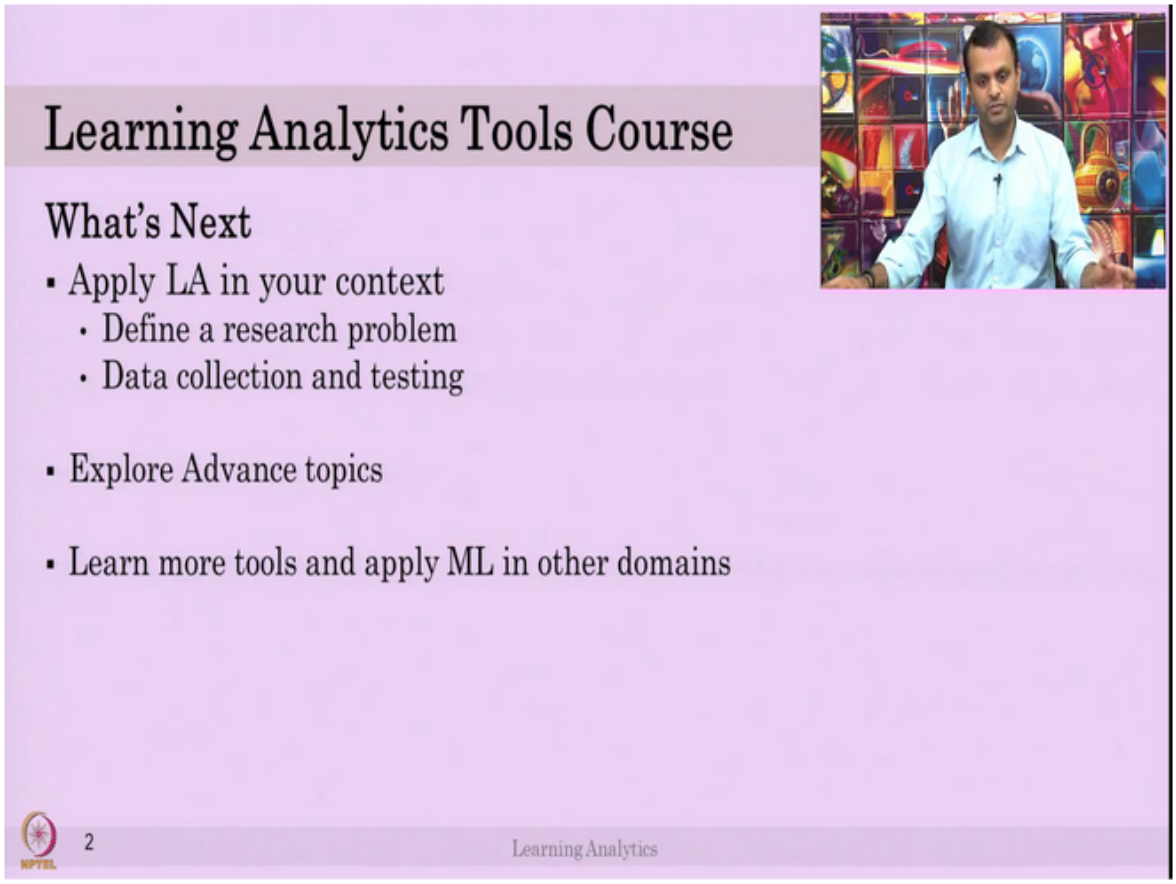
Indian Institute of Technology, Bombay

Lecture No. 60

Source of Data collection and Research Community

Welcome back. In this video we will talk about data collection i.e. how do you collect data, when there is no data and what is the research community. So, after learning this course what is next, so next is you can apply learning analytics in your context, or if you are a teacher please apply it, if you are a student you can apply also learning analytics if you collect data, and then you do the testing that is the next thing to do.

(Refer Slide Time: 0:53)



And if your interest is towards doing research in LA, collect data for different research questions and do testing and apply it and write and publish papers, that is what I want, I would be happy if you do that.

But if you are not a researcher and have just joined this course to know-

1. What are the basics ML algorithms?
2. What data is available in the educational domain?

so to go and explore advanced topics then (as I was telling in the last video) you can go to the last video in which we had talked about how you can find resources on the internet and read it.

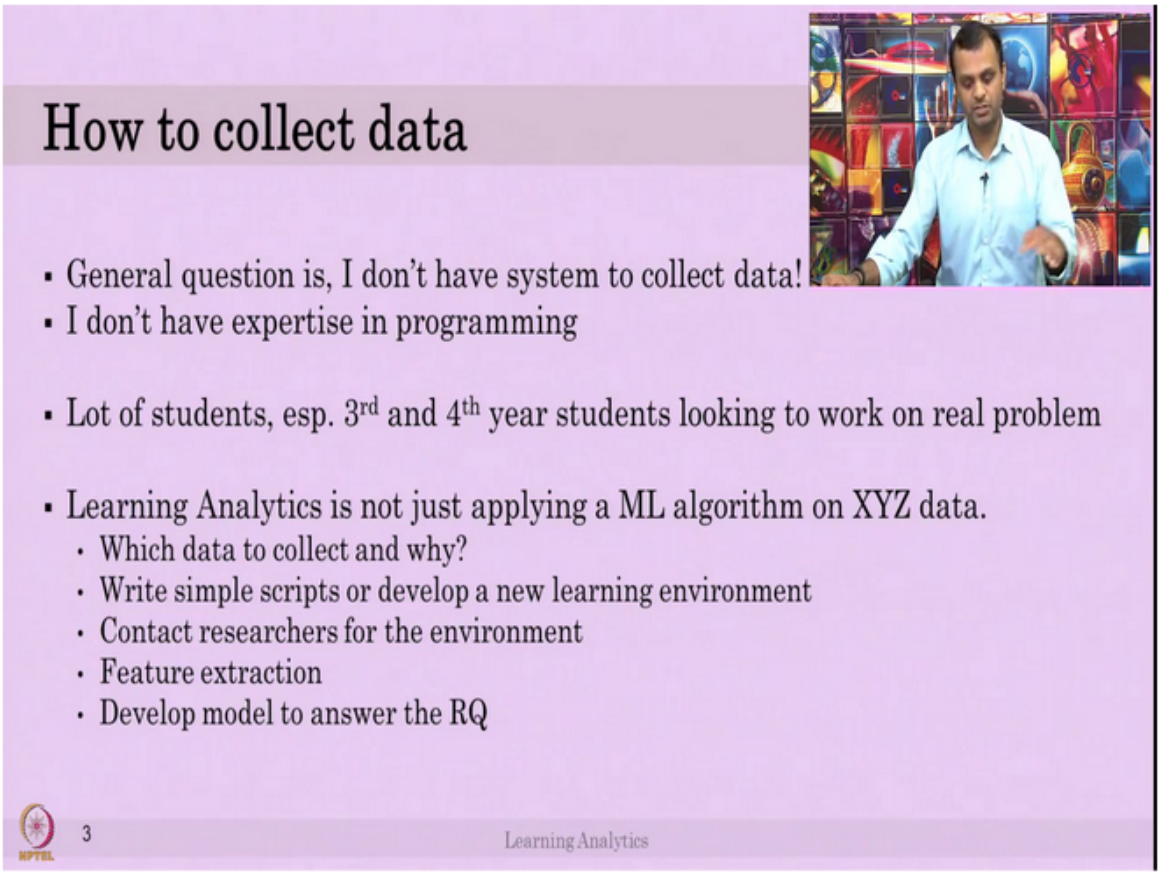
If you are interested in the tools or data collection and you are looking at a job in entry-level jobs in a data science career, I would recommend you to go and learn more tools, apply ML in different domains and create a repository where you have applied data, you created models and

maybe create a GitHub library or create a web page where you applied whatever you learned and you then apply for a job, then it might be easier to get a job, so that is the whole idea of three directions you can go.

The 3 directions are -

1. I want to just know more about machine learning.
2. I want to do research
3. You want to look for a data science job, do not just stop with our learning domain, apply for other domains or more about tools it is all interesting. I will talk about what tools are used in the industry and everything about how to find it out, I will talk about that in the next video. In this video, I will talk about research kind of flavour in LA.

(Refer Slide Time: 2:18)



So, the general question everybody asks is I do not have data, I do not have a system, I am not an expert in writing scripts, I do not know data, which data to collect, where to get data from, there are plenty of data already available in online like a kaggle or DataShop everything is available but that will not answer the research question you are looking for if you are trying to answer a particular question I would recommend you create your own system and collect data.

But not everybody is a computer science engineer, I do not have a system to collect data that is perfectly all right. The only thing is due to advances in the computer science field there is no need for you to code or know programming to create the environment, the things are changing. On the other end and I see people, a lot of students an especially third year and fourth-year students, third years, or final year students in different departments know engineering or science, they always come and ask you looking for I want to work in a real problem, computer science, mechanical, civil, electronic does not matter which department, you know visual communication, they all want to learn to program and they want to create something.

So, if you have access to students, give this as a challenge, give this as a work, ask the students to create a system, ask the students to create a front-end, back-end system, maybe Django with Python or Node.js, Ruby on Rails something like that, ask the students to create a system where server-client communication is already established. And ask the students to observe all the clickstream data in the client store in MongoDB or any structured SQL, the students understand all these things and ask them to create. The moment you ask them to create this kind of data they will be very happy to do it because it is a real-world problem.

I see a lot of students come to me saying that I want data, I want to do something in machine learning, I want to do, and they look for from me is a data I said no I do not give data to anyone, the reason is, it is the data I collected based on my understanding and I can give some data but it does not serve your purpose. If I give data to you and you click something a few clicks in the machine learning tool and you create something that is not going to be any learning. Anyone can do that we want you to develop a complete understanding do not need a complete understanding.

For research, you have to come up with your own research question and what data to collect and how to use the data to find the answer to your question. So, look for people, students especially

who are in third and fourth-year students who are interested to work on a real problem give them this as a problem and they will help you.

If you are a student already you are looking for it, take this as a challenge and create a system, check for Django python or Node.js or anything which is latest and recent and try to create a system where server-client communication happens, all the client's clickstream data like watching the video, reading page, everything clicks has been copied, moved to the server located then test it with your network and if it works upload it in some digital clouds, there are free clouds available for students, for the student id they will have some free clouds available use those services, upload it and conduct the study, this is how you have to start.

If you say I want to start only with the data and extract features then that is not really a good job, if you say you have created a system, collected data, answered this question, the prospect and your job market also increase, if you are a researcher you want to look for this data you are creating a new system, you are contributing something new to the community saying that this is our system and your name comes up, there is a system created and data is collected and there is a study done. So, you are establishing a research network within using your system.

So, I always recommend going for that. But no I do not care about that I just want to apply ML algorithm on XYZ data that is not learning analytics especially I was thinking, you need to know what data to collect but you do not want that I just want to do that, I would say go-to online resources.

Learning analytics is more about which data to collect why and write the simple scripts, in the very simple way what you can do, you can create a Google site it is easy to create, there is no programming knowledge needed to create a Google site, you create multiple, you create multiple pages and you can add videos and you can ask the students to go and take a quiz and you can write a simple small script in the Google which captures all the student's clicks, I would recommend to go and check that, how to do that, ask the students to do it, it is all you can start with that.

So, if you are interested in one particular environment say I have just looked at this paper this environment is available I want to use so it might contact the authors they might be happy to

share, some will be happy to share if it requires a lot of training and onboarding process they will not do it, but if it is available freely online, talk to the researcher, researchers will be happy to share as long as you acknowledge them and you inform them what are you doing with that.

So, contact those authors they will be happy to share, but make sure that what you want and what is the data and talk to them in detail, simply I want the system will not help, you have to give a detail why, why did you read that, what you are planning to do, give those interesting kind of reasons for some synthesis on that interest to that particular environment which you read in a particular research paper, go and talk to the authors, that is the one way you can do, I actually do that with a couple of authors who created environments because I do not myself create environments as such as of now.

And more about after that you should know what is feature extraction and models from the research questions and solving it is what learning analytics is, it is not just applying ML on XYZ data.

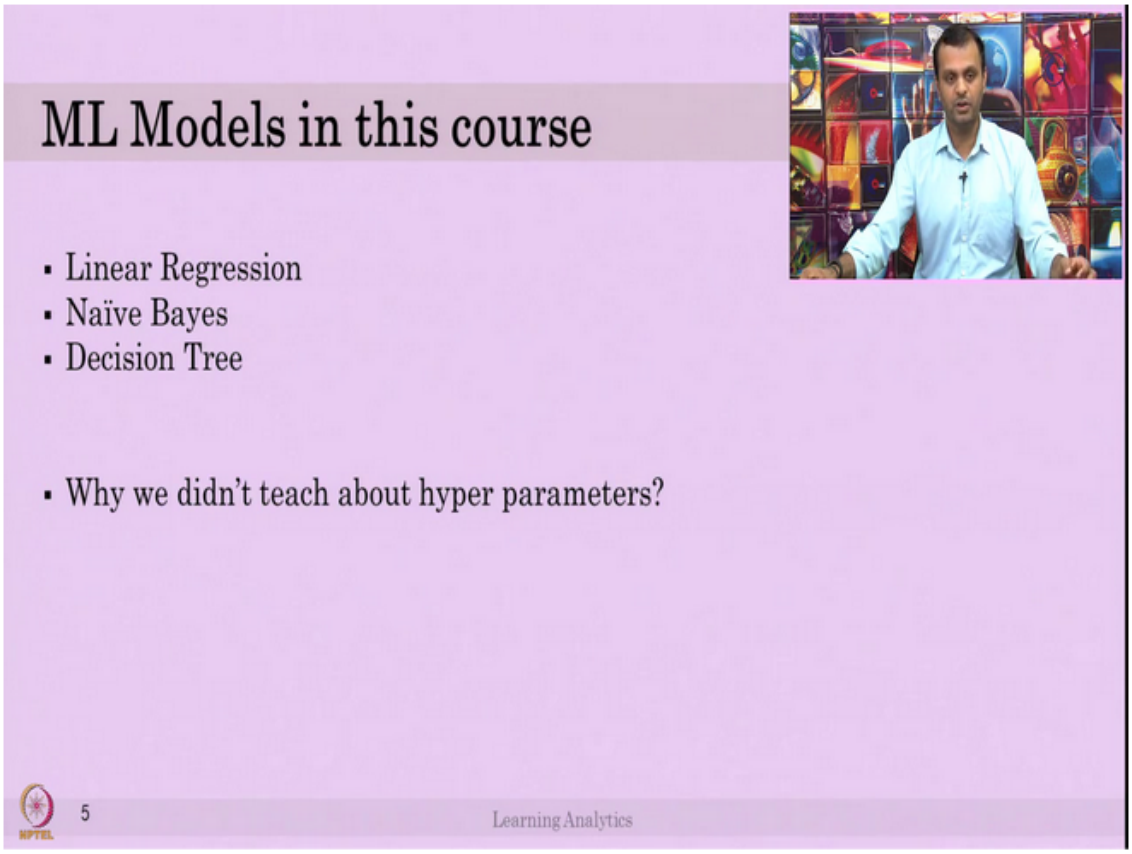
(Refer Slide Time: 8:19)

So, if you do not care about anything and just wanted data, go to DataShop up and straight learning centre created something called data shop, just hosted in a CMU site, go there, look at the data, register yourself, you have to contribute a data but fine if not contributing you can access to a basics data look at the data shop, a lot of data from all the resources, most of the resource in academic learning environments uploaded that, can use it. And the thing is the data available to you that you will not get more data if you want to do another kind of research you will not be able to do it but with the data, you have to live with that and you have to answer such questions on that.

Another good resource is Kaggle. Kaggle now and then upload hundred, thousands of users data, clickstreams, a lot of data has been uploaded in Kaggle and that Kaggle data is also useful, you can do and download the data, you can ask a different research question from the what has asked in the Kaggle, so do that. But Kaggle may not allow all the data can be freely used and for such purpose, so check that the constraints and talk to the developer if you are interested in analyzing

the data for some other thing, something else, get the permission, that will be needed for the publications, if publisher finds out this data is not diverse and is available on Kaggle and they will check whether you have got permission from Kaggle or not.

(Refer Slide Time: 9:49)



We talked about in the last video also, we did not talk about Linear Regression, Naive Bayes, Decision Tree, these 3 I did not talk about in a detail. So, I again repeat that go and watch the videos, I request you to watch Andrew Ng's video to understand what are these three things, if not any resource talks about machine learning and hyperparameters, books or videos will be good.

(Refer Slide Time: 10:16)

And now come to the research areas in LA, LA is just a topic, what I was trying to say is I collect data or use machine learning, what is the areas I want to do? The basic thing everybody does is modelling learners performance, modelling learners engagement, modelling learners skill in the environment, modelling learners to predict their performance in the end sem (marks or something like that). Or engagement is something new and it is been for some time but people are working newly on engagement to say how multiple environments are contributing to the engagement of student.

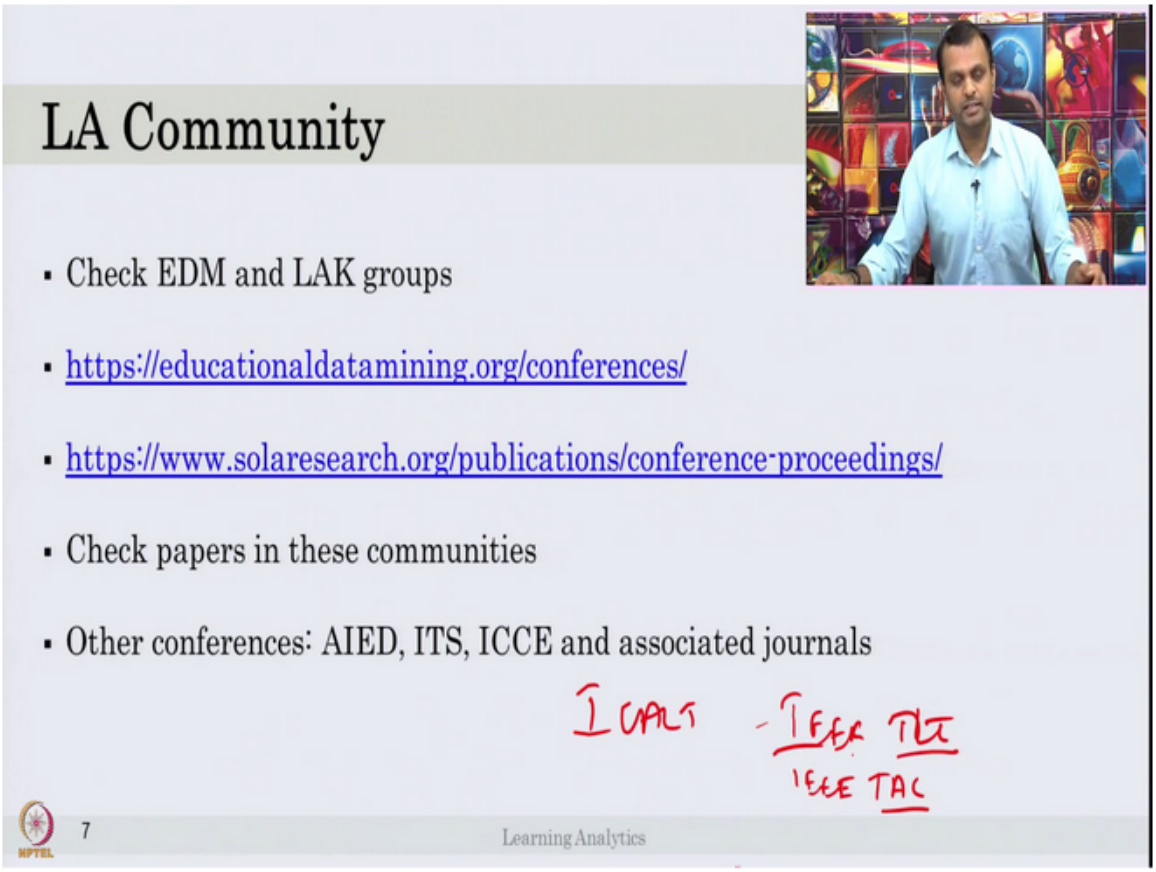
Or content analytics, education is all about content, you are giving a video, you are giving the test or other contents, so a lot of people do on content analytics just check the content. Or you can apply machine learning and you can apply pattern mining algorithms and you extract the strategies based on students interaction data, what is the strategic difference between high performers versus low performer like a DSM differential sequence mining, you can do that.

A lot of people do research on privacy, how to store data, what is ethics and privacy also, there are a lot of research papers published on those topics, look at those papers. I talked about affective computing in a multi-modal learning analytics introduction, so you can do the multimodal. People also do a data visualization like a dashboard they innovatively come up with the new things, how it is useful for something you can, you saw the iSAT, it is kind of a new direction in the data visualization in the learning analytics community.

And the last decade people talk about self-regulated learning and its advanced topic, it is not just one data, data from multiple channels and multimodality of data and you have to analyze that to create a self-regulated learning model. And the advanced topic is intelligent tutoring systems or personalized adaptive learning environments where you not only predict it and you are prescribing them because just suggesting something new to students to achieve something.

And a true intelligent learning environment which should be completely free from any fixed set of rules not exist in the world, lot of people do ITS intelligent tutoring system just based on the performance in the current exam, not even the history of performance, the successful ideas all do only the performance-based adaptation, so you can do something in that angle also.

(Refer Slide Time: 12:49)



How do we know about these research areas, how do we get it I will tell you. What I did, I went and checked the EDM education data mining and LAK groups, these are the conference proceedings available, this is the EDM, this is the EDM conference resource, you can join check it out and look at the abstract papers and all this EDM conference papers free for everyone, look at those abstract papers and check what is the title from the data you understand what is the topic I was talking about.

Similarly, for LAK they put all the things in the solar research community what I would say is to go and look at the paper if you do not have access to the paper, check the title in the Google scholar and see Google scholar might give you the PDF access, and if you do not have further look at the similar papers with the same authors which have access to you, nowadays most researchers, most authors upload their papers in their own personal web page, go and check the personal page or the research kit it is all happening, people are moving towards more open compared to keeping it for the cost.

There is no need to pay to a journal unless it is necessarily important what happens is you can talk to a researcher and you can find out similar research publications by the same researcher it would be available go and read that, educational keep completely open-source, so yeah you can get a lot of open-source data. Check papers in these communities.

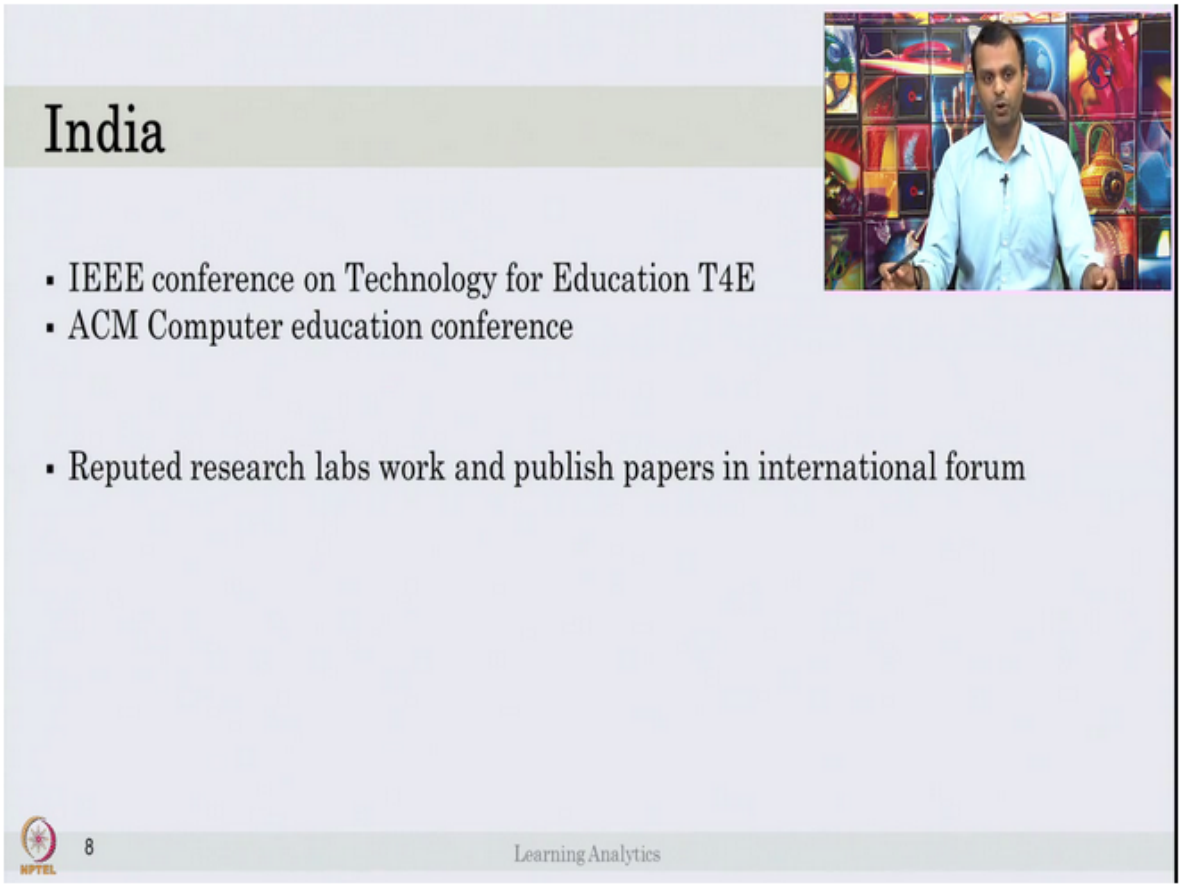
Other conferences which are really interesting is AIED artificial intelligence in education, it happens every year now, so look at those conference or intelligent tutoring systems this also does something in learning analytics, modelling students to provide something. These are international community EDM, LAK this, and there is something called so ASIA specific things called for Asian international conference on computers in education, this is by a group SIGCHI in Asia and so check this not APSCE, so check this website also international conference in computer, international conference on computers and education.

And each of these guys now has their own associated general IJED, RPTEL, JEDM journal of EDM or LAK journal or there are something called ICALT, there is a good conference called ICALT this also international they have a conference on IEEE, transactions called TLT, or IEEE transactions on affective computing, TAC (transactions on affective computing), transactions on learning technology, a lot of journals associated journals, when you look at this conference you might see get the associated journal names.

Look at those papers and get the research field, if you are looking for which research topic to start in LA or if you are looking for which topic to start in LA, I would request you to go and read the last 2 years conference proceedings of EDM and LAK, not all proceedings just look at the title and abstract, if you by reading it you will come to know where is your interest, last 3 years, 2018, 19, 20, three years is a good time to start.

And look at it last 3 years conference abstract and proceedings you will understand what is the topic is about, then pick the topic which you are interested in and topic which you can do research on, which topics which you can collect data, we have access to those resources depending on that choose the topic, these are some other good interesting conferences where you can publish your work, so that is about LA community.

(Refer Slide Time: 16:36)



In India we have, we run this conference called IEEE conference on technology for education, it is mostly technology but we also do a separate stream for learning analytics in it, so if you are interested in Indian conference look for this T4E and check those papers it is all open access after one year, so check those papers. Also, ACM computer, ACM group in India conduct the conference on computer education to check those papers also.

And this is a lot of reputed research labs in India for example IBM or Wipro or Adobe they publish papers on the international forum, it is not that there is no education research or learning analytics research not happening in India, there are a lot of good companies for example education initiatives do research in this area, all the commercial companies also do research in this area and publish papers in the international forum. So, look at those companies' articles if it is available online. Thank you.