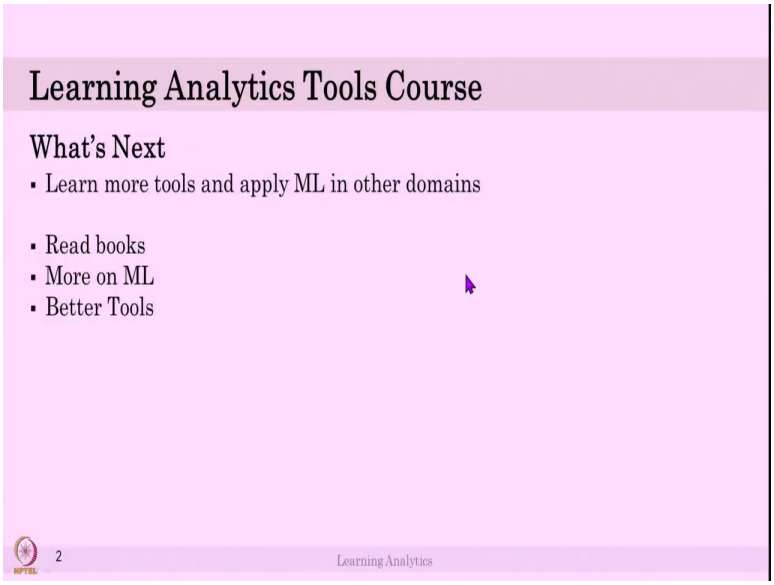
Learning Analytics Tools
Professor. Ramkumar Rajendran
Department of Educational Technology
Indian Institute of Technology, Bombay
Lecture No. 61
Machine Learning tools used in industry

Welcome back in this video we will talk about what are the tools used in ML infrastructure in the companies.

(Refer Slide Time: 0:27)



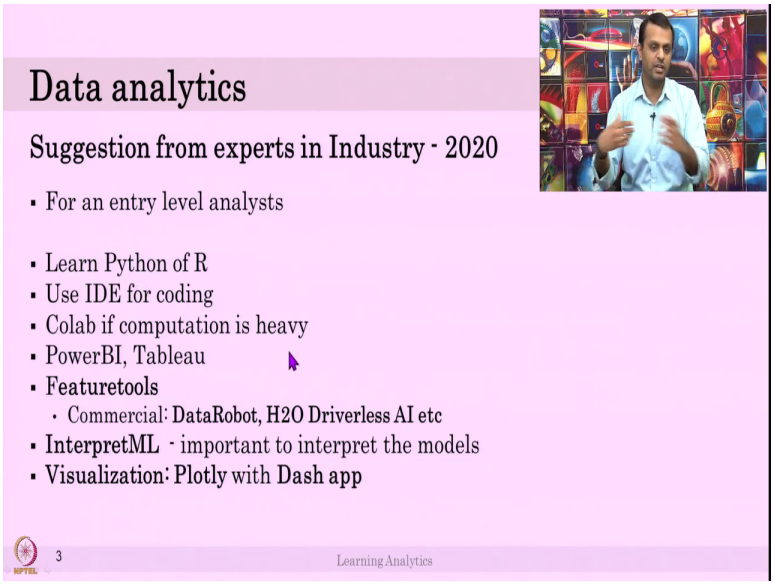
So, if you are interested in the algorithms in ML domains or you want to join a job at a scientist level, the data scientist industry, so we will look at what is needed. So, I would suggest you start reading more books, there are very good books available on machine learning.

You are worried about what are the best books to read, do not worry, go and check it on the internet there are plenty of advice for you and look at the suggested books, pick the second or third book they suggested, it always works well. Some of the well-known

books like professor Andrew Ng wrote a book from his course on machine learning or Christopher Bishop wrote a book on pattern recognition in ML.

So, it is based on what is your interest, if you are interested in learning machine learning in a programming language, script languages like Python or R then choose the books, if you are interested more about mathematical aspects how it is happening to choose books which serve that purpose. Also learn a lot about better tools, do not stop at Orange or Weka they are not enough if you want to give a job, so more tools will be good.

(Refer Slide Time: 1:48)



So, based on talking to my friend who is a data analyst in a multinational company abroad, I asked him what are the tools you have to use, what are the tools they have to learn for entry-level analyst in machine learning, what will help students to get a job in the industry? This is in 2020. So, he said that learn Python or R, it is very important, it is a scripting language, everybody is moving towards Python or R, and if you have good skill on C++, C sharp, Java, still works but Python is what is taking up.

Use IDE for coding, some interacting development environment, an environment which supports Python or R is what is suggested, because that helps a lot to learn to code if you do not have resource just use the Google online resource, Google Colab. That is it, if you have heavy computation so you want to apply machine learning like neural networks on millions of data, you have multiple layers, you do not have a system with you, you can upload the data on Google Colab and do it, that is very good. The only thing is to make sure that you are uploading data, the data have a highly identified anonymous, you do not give the data to Google, that is important.

So, the tools he suggested go beyond Orange, or Weka or Rapid Miner, go for power BI, Tableau kind of tools and these tools help you to understand work with large big data kind of thing. The tools we use are good for the small low size data, in the industry they use data from millions of customers, and they churn out this data to predict something good. So, use the tools like Power BI, and Tableau will be helpful.

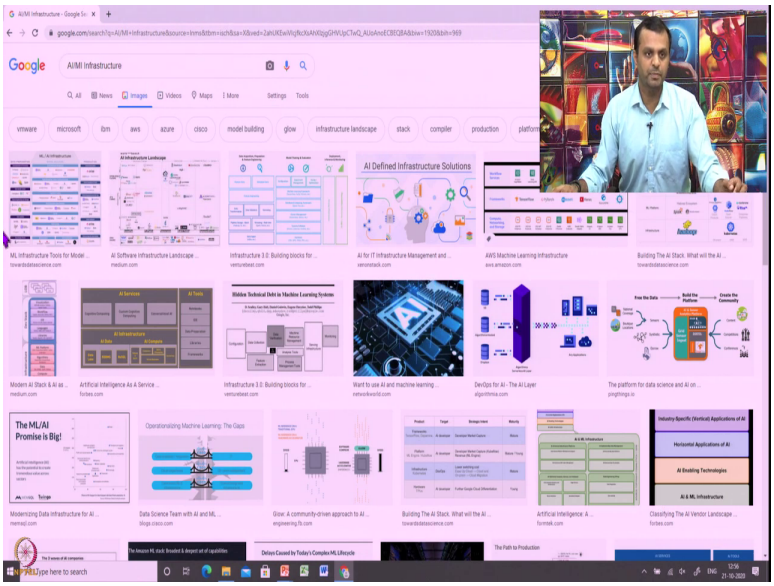
How to learn that? Most of these tools are available for academic purpose for free but we might get a free licensed version for some restricted period of time. There are a lot of videos available on how to use Tableau or Power BI on YouTube. For features, extraction is also we have some tools like DataRobot, H2O, Driverless AI etc but the point is it is all commercial. For the noncommercial purpose, we talked about one feature tool for feature extraction in week two, we gave one particular tool.

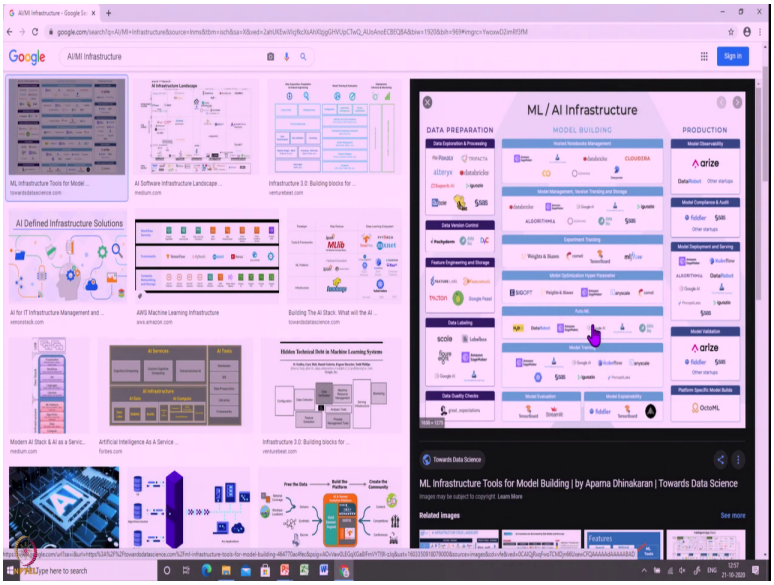
So, there are noncommercial open source tools also available. If you do not find, there is a commercial, I do not know what it is alternative, go to a website called alternative to, or type alternative to DataRobot in Google that website will come up in first, it will suggest all the other open-source alternative, they give the detail to you, so pick those tool and learn more about data.

The more important thing in the industry is not just collect data or extract features and apply some ML-tool, beyond that are interpretations, so how do you interpret the models, so they all go towards decision trees, or linear regressions to interpret the model, explainable AI.

So, use the tools called interpret ML, it is very very important to interpret the models, and they help you to do that. So, these are the tool's suggestions from my friend in the industry, he said that this is useful if you are looking for a job in a data science carrier. And he suggested visualization Plotly in Python, it has a dash app, dash app to generate via developer university which creates interactive dashboards, this Plotly actually has this dash app also included so it is all in private then, so if you want to create interactive things it is all easily available there.

(Refer Slide Time: 6:05)

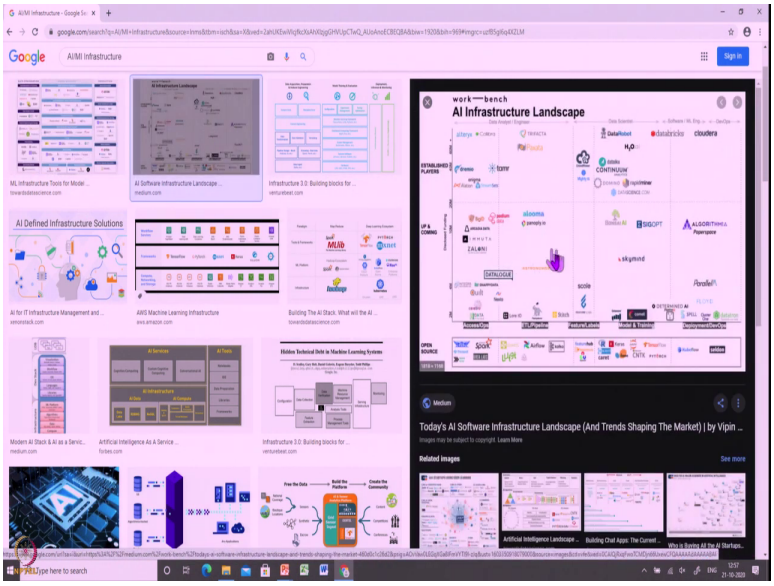


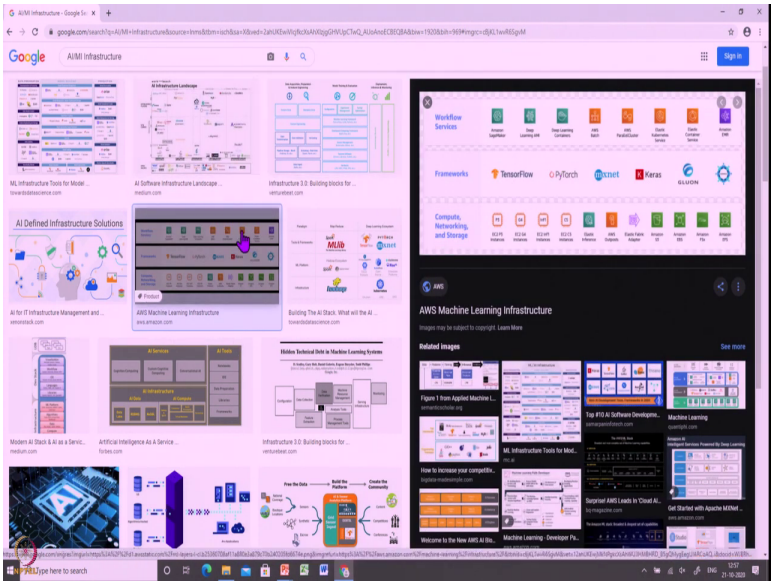


So, if you do not know these tools or stuff, this is one person's view and is it true or not, if you want to know what is the current thing in the AI/ML industry I would say this, go to type this words like type things called AI or ML infrastructure, that is it. We type it, we look at the first one, let us see that detail. See there is data preparation, model building, protection, you might want to look at the website and they will explain each and everything and they will explain why it is used.

I was telling about, is it too much big data you can use this, you solve data preparation explanation process and feature labs or feature tools, the feature tools is going to be an open-source one, just use it. And in model building you can use multiple ways like tensor flow or different model builders, this is how the things available, so go and look at the current things.

(Refer Slide Time: 7:14)





And this is always good, you can go and look for latest things from Google. So, what I am saying is if you are really interested to take a carrier in data science, just look for these tools which are open source and learn them, not all of them there is nobody who is going to be, nobody is going to learn everything in this. Pick one tool in each section and learn about them if you have time pick one or two more tools in each step that is data preparation and modelling and interpretation.

And integrating the tool with the industry is important, it is not that interpreting, how this is coming back to the industry, how they are giving feedback to the machine. Learn about this one or tools in each step, pick it up and create your resume and apply data from Kaggle or any online resources, pick it up. This will then you apply the different domains either use data from different domains, this will give you a good set of resume to get hired from the good companies like data science companies, that is the idea.

(Refer Slide Time: 8:25)

So, I suggested some tools based on a friend's suggestions or you can go and check online for the latest tools. So, thank you for taking this course, I hope you really enjoyed and you learnt something and thanks for being with us for last 12 weeks and thanks for

doing all the assignments, I hope you learn some tools and you take something from this course like I want to do some research in LA or I want to learn something in ML, whatever direction you take, we are very happy for that, and if you have anything further questions or something, talk to your peers who are learners in the discussion forum ask questions as the learners to respond or discuss with the learners or do reach us if you have any questions, special questions how to do that.

As I mentioned most of the questions can be answered by Google already there is nothing else that we know better than what the world knows. So, but keep the learning and I hope you really enjoy the course. Thank you.