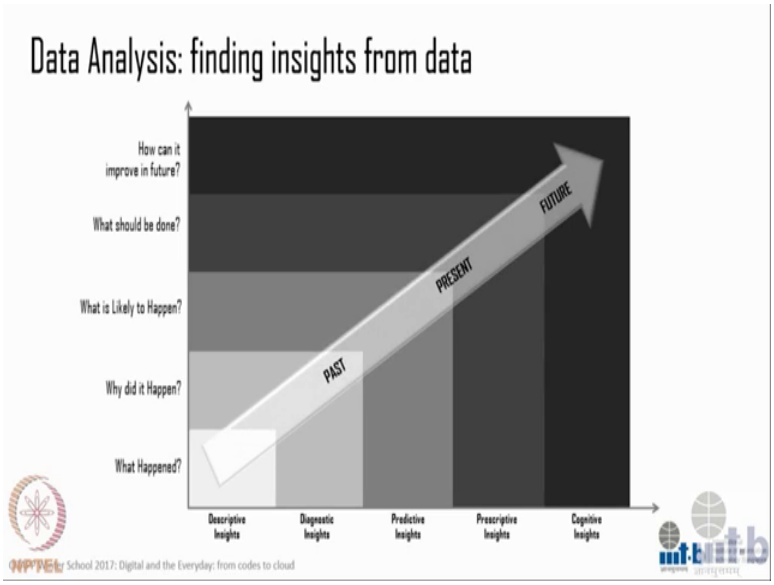
Digital And The Everyday: From Codes To Cloud
Prof. Bidisha Chaudhuri
Department of Multidisciplinary
International Institute of Information Technology, Bangalore

Lecture – 10
Data Driven Identities Part 02

(Refer Slide Time: 00:17)



So, what happen? Or why we are so interested in data? Because of this the tools of analysis that are available at our disposal, right. That is something makes data so attractive. So, if we are moved on. So, all these evaluation that I am talking about, right of the digital data, and also the platforms and the data storage capacities that has evolved over time, it has moved on like this. So, this is the y axis is basically a time the temporal trajectory of the analysis. And this is what the analysis tells you. So, I will give you first one is the descriptive insight right.

So, if I ask you; so, what is it descriptive insight, if ask you how many women, or what is the percentage of women in India in engineering education, right. That is a descriptive insight the data gives you. Now from that, if I want to diagnostic insight is I want to find out why, right. I want to find out an explanations that let us say why only 2 percent of engineering students in India are women. So, if I want to find an explanation of that, if I want why did it happened, and that is the diagnostic insight. You can all get this PPT's later.

So, it is. So now, how do you that? That is when you need this relationships that we are talking about between different columns. So, it is not just your gender and engineering education column, you need to understand where do you come from, right. What is your age? What is your family background? All these relationships needs to be created for me to be able to explain why this happened, right.

Now, the next is predictive. So, predictive is, when I say that given my diagnostic insights, given my understanding of the relationships of the demographics of women who come to engineering education, I predict that this segment of women, right. This particular combination of demographics of women will are likely to come into engineering education.

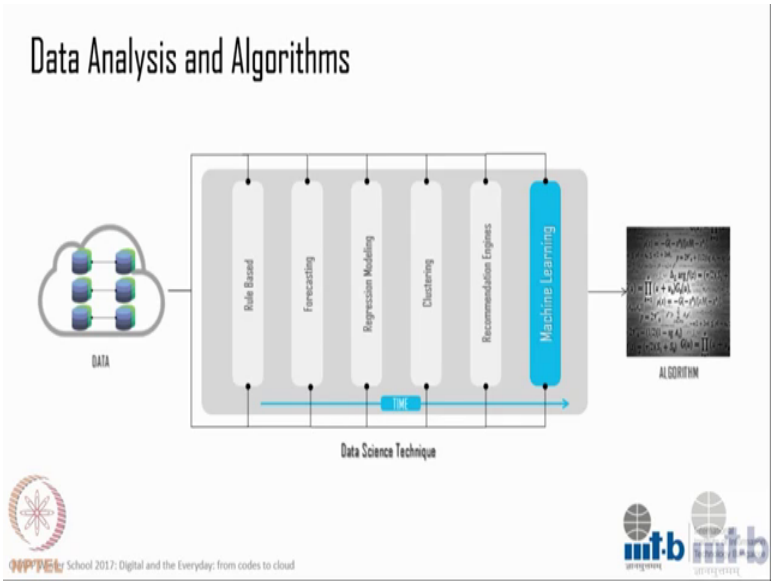
So, let us say I can predict on the basis of this diagnostic insight that by 20 20 so many percentage of women will be in engineering education in India, right. then come prescriptive. This is where we are pretty much right now. This is what your recommending recommended system works on. This is what you do why you do. So, relating it to other things that you do and. So, you are likely to read this. So, I prescribe this to you. Does it make sense?

So, but this is not the end. We are moving to something called cognitive insight. This is what AI, machine learning, all these things are going to do. So, from descriptive to prescriptive will all be done by the machine. We do not need to do it anymore. That is what we are headed towards with machine learning and AI. Now the implications for this. So, I want say it is an implication, but just to give a sense of what it means, when we move from prescriptive to cognitive.

So, let us say we were talking about algorithms in the you know the couple of session back, right. So, when we are talking about algorithms our algorithms will be able to say that if these are women who are not coming to higher education, the reason they are not coming to higher education is where English education is not good enough.

So, let us focus on English education, that is a prescriptive one. And on the basis of that people who are not following that prescriptions, they will be out of it. So, you will never look at them as your probable you know what you call them (Refer Time: 04:38) or the sample who will be likely to be in engineering educations, right. is this slide clear? Ok.

(Refer Slide Time: 04:53)



So now how this is all happening? Some try to do connect the data platforms that we talked about, and the algorithms that you know are running. So, these are data techniques. So, it was first I mean I am sure people who have being doing this for a long time would know this better then me, is this rule base 2 machine learning algorithms, that we have moved on to right. So, these are the techniques through which now the data that is being generated, that all the kind different kinds of data that we were talking about, with the help of the data platforms, right that we must just elaborated on.

We are running certain algorithms through those data, and your are able to create insights. And in terms of the techniques that we are using, right. To run this algorithms on the data has reached where this is the current state we are in. So, anybody wants to explain how machine learning algorithms are different or what is the unique about machine learning algorithm, Shravya?

Student: Different from all.

So, all that we were doing till we have reached till the recommended system, right. That works I mean that we pretty much experience on a daily basis.

Student: (Refer Time: 06:16).

Right, but we are now trying to move on to the machine learning, that we have as an users as an end users we have not experienced it yet. So, if when it comes a reality, what it is going to look like?

Student: So, some of the models that are that also have machine learning, but the reason data can machine learning techniques, they have a kind of complicated models which it draw inside a (Refer Time: 06:36).

Student: And these are extremely hard to explain. And they have a much strong capacity draw a insights draw a (Refer Time: 06:40).

Student: And it is like it is more like they have become extremely complicated themself.

You want to talk about the neural network a bit.

Student: (Refer Time: 07:02) To what context and all?

So, I am saying that see for example, when I used to the Facebook or let us say; so, I have been reading. So, let us say I read guardian a New York Times through by Facebook, right. So, all there is articles that are published on this 2 websites, always come into my newsfeed on the Facebook. That is also an algorithm, right. And that is where the recommendation engines works, right. That is the that is where the analysis. I am saying when we have machine learning algorithms doing this jobs. How it is going to change my experience.

Student: So, the machine learning algorithms can also produce definitions.

Student: Thing is that particular implying algorithms, they are much. So, there is limitation computation limitation to bring this algorithms.

Student: we can only produce this limited amount collection we can draw limited amount of connection between data and (Refer Time: 07:54). So, the more hardware connections can be drawn with current machine learning techniques so.

Yes.

Student: So, it is more powerful in (Refer Time: 08:00) that long (Refer Time: 08:03) like, how I am my family is similar to somebody else. How I am going to my (Refer

Time: 08:08) should take me like (Refer Time: 08:10) social media is working things like that. So, these could generally take up small small collections and timely or several connections, but we (Refer Time: 08:16) and pretty accurate connections right.

Um and it also I think what would be an interesting part of machine learning algorithms is, the self-corrections right. So, it can learn.

Student: Regarding (Refer Time: 08:28).

On the fly.

Student: That supposed to be the.

That is.

Student: provide internet to everyone.

Yes. So?

Student: the thing is definitely we have is like, we have a different multi layered concept.

And deep learning.

So, we have different way the connection between different things, it depends on the weights that there have been assigned.

So, as and when we feed in more data to the system, the at the value of these weights between connection changes. So, that is basically how recommendations and a suggestions are provided. So, the higher the weight that will be; so, it will be likely probable that that has more relevant to you.

So, and also it establish the relationships. like, a higher the weight between 2 connections; so, that will be like a of more significance.

Student: So, it provides us a relationship between all of these.

So, just to you know collate all of this that has been said. So, as Shaadir righty mentioned that the machine learning does all of these right. So, it is not one is different from the other, it is a journey. So, of course, this one is able to all this things at the

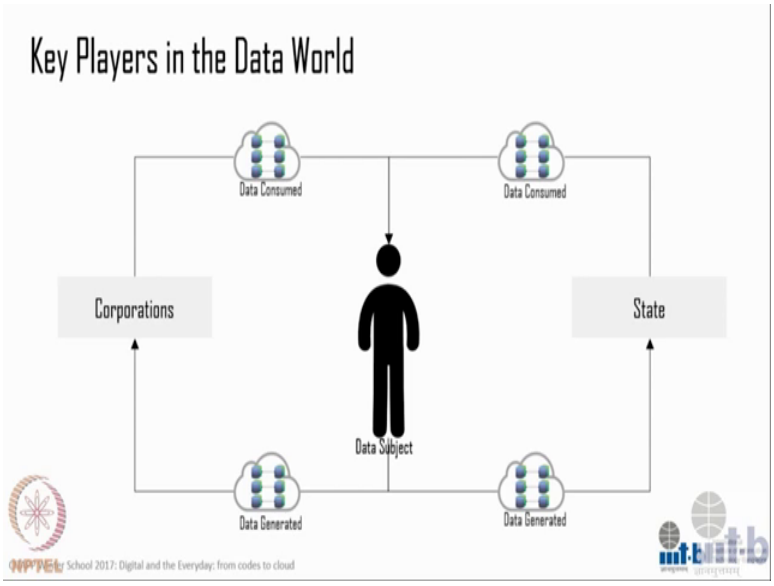
moment so, but when we went to move to machine learning, and it addition to all that is before that, what it can do to put it in again in a very you know more general term, is that it can act better than a human cognition. That is what artificial intelligence are headed towards.

That you can do all these much more efficiently and much more complicated fashion. So, you are trying to build a more complicated machine than the human brain; which is modelled after a human brain, that is what neural networks talks about, right. That is the network that our brains operate on, but it is trying to do so that all this can be somehow automated, right. And what it can also do that if it can learn from it is previous experiences.

So, let us say I have drawn some prescriptive insights from the datasets that I have been working on, but if it is not working, I am able to change, I am able to correct the algorithm and try and fit in the outliers. That was not possible, possible in the previous algorithms. So, the previous technique attributes clear.

Now, how are key players?

(Refer Slide Time: 11:00)



So now we have this is the material world in which the data operates, right. so, we talked about the kinds of data that it exist, the types of data that it exists, and how data stored what are the platforms that exist, how we draw insights from that data which makes data

so attractive and also the you know, the analysis that are you know run on this data. So, this is what you here when these days, right. These are the buzz worlds of current times machine learning algorithms AI, cognitions, neural networks clouds all this things, right.

So now how all these things are taking place? So, what we are saying? So, this is these are data subject these are we. There are 2 major place. I am not saying that just 2 place in the sense of numbers of players, but these are 2 segments, the major segments of players. One is the cooperate, the cooperations big cooperations how are interested in this and the state. So, the further corporate; so, what we are talking about, that we are generating this data, right. Now this data is somehow running through this algorithms, which are mostly preparatory right. So, I do not know what algorithm Facebook runs. That is their IP, that is their intellectual property ok.

So, it runs through this algorithms. And then new data is generated, which I am again consuming, clear? On the other side, when the see the state, who are equally interested. So, for the cooperates it is a so did this you must have heard this like data is the digital currency right. So, by running those algorithms, that we are just talking about, we are able to generate these insights; which makes the data that digital currency.

That is what you are selling insight, right. Because you are able not just use this techniques, and the volume the storage capability that we have along with all this machine learning techniques, we are able to generate insight, we are able to mine data in an unprecedented volume, right. And we are able to draw insights; which are of extremely sophisticated nature. That is what make data a very profitable commodity.

Student: (Refer Time: 13:40).

Right? Now why state is interested in it? So, what the state is interested in the data? For that it knows more the subjects, knows more about the subjects, right. It subjects, and then it can use it, that is you must have heard the term evidence based policy making.

Student: ya.

Right? So, this is what the state is interested in. So, it knows not only about let us say your gender and your educational qualification, it knows what is your bank you know financial data is like, how you do transaction on the financial space, what movies you

watch, what e commerce platform that you use, what sort of social network that you maintain, what is your political opinion, all this things right. So, your consumer behavior, your political ideology, your social behavior, everything is the tip of a data.

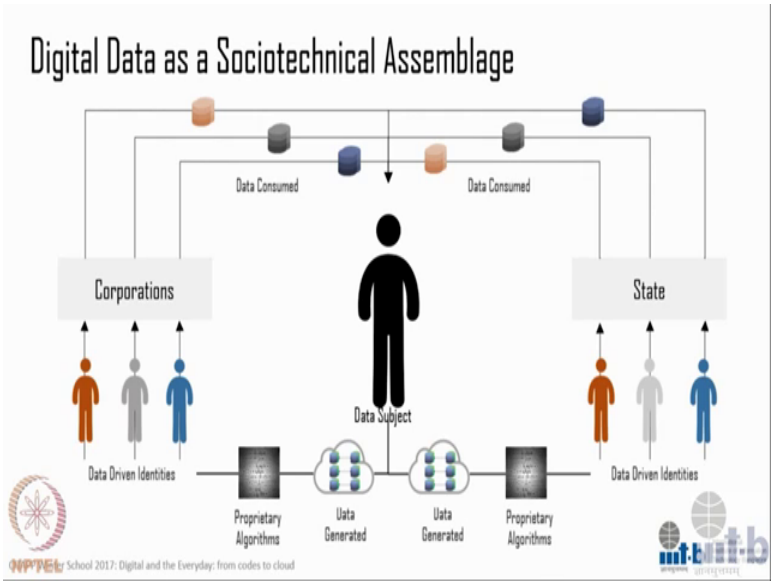
Student: and a constituency.

Right? So, state of course, is interested in this data. So, there also I mean not so much may be, the Indian state may be AADHAAR would be an example of that. But in us this is already there are examples where the state also runs this proprietary algorithms, on the data it you know collects on it is citizens, to make better decisions better policies, and also laws and regulations, right.

So, this is pretty much how the key place. So, looking at us as the data subject, and then looking at how we are sort of governed both by the cooperates and the state as a data subject, right. Now this is what the material aspect of data. So, that how the technically speaking it works; like all these things that you buzz that you here about, this is what it pretty much it sums up to of course, it is it is a just a session you can. So, a whole course on this. But I just want to stop here, and ask is there any clarification that you need, no? Ok so, we move on.

Now, we are trying at that the other access that we talked about, the material and the social. So, let us see what is the social implications of all this material things that are taking place. And of course, when we mean social there if can be many things that we can talk about. But in for this session I have just focused on this notion of identity, right?

(Refer Slide Time: 16:24)



So, we were talking sociotechnical assemblage, right. Yesterday which basically means that technologies do takes place where it is a gathering of different actors, and this different actors can be artifacts, can be text, can be people, can be organizations, can be you know languages, all this things that we were talking about.

Now, how can we understand digital data as a sociotechnical assemblage. So, this is again I am summing up what I just said, right. So, of course, then there is this data subject, this is the data I am generating which is run on a proprietary algorithm; which is creating certain kinds of data doubles. That is the word usually used. So, basically let us say if I am going to buy ticket from make by trip, that is my travel identity. So, make my trip would know you would get offers from make my trip for certain, let us say, if you go a lot to Delhi right. So, most of your offers would be you know around that.

So, that is your travel identity, then if you use a Swiggy or Zomato, right. That is your consumer identity in a very specific segment right. So, all this and then you use social media data. So, all these data that you are generating passing through this algorithms, and creating certain identities of yours; that is, you know a profit proposition for the corporations, and accordingly new data that is being channeled to your way, right.

And on the other side it is pretty much running on the same way. It is just the purpose of doing this is the slightly different. So, there it is mostly trying to look at you as a consumer, and the state is looking at you more as a citizen, right. Or the subject to be

governed. Now I want to focus a little bit on this data driven identities that we are talking about here.

Student: So, why does for example, let us say, at the a university as an example comes (Refer Time: 18:43) it is not strictly, it is (Refer Time: 18:46) just talk about it suppose, but it is not up here cooperate.

Student: Enterprise.

Student: but at the same time, we have to love to get that (Refer Time: 18:54) on some data. So, we can take the course (Refer Time: 18:58) where 2 to 3 fog members will be questioned, and just as an example.

So, why I as I said the. So, it is not like I am not trying to exhaust the number of place here, I just clubbed them on the basis of the motivations.

For which they are interested in the data. So, let us say if I am trying to govern my students with the data. So, it not just about selling my course right. So, it can be a it can has a have a profit motive. It can also can be there, I want to give them a better experience when they come here, right. So, and of course, if I break this things down, there will be overlaps.

Right? So, state and the cooperates in this case work pretty much in a nexus. So, even sometimes motive wise you cannot really differentiate right. So, I think it is a very good question you asked at this point. So, this is just to give structured what I am talking about, I have tried to club them on the basis of their motivations. Because we were talking about that in a sociotechnical assemblage, the different actors have different motivations. Sometimes this motivations are not in you know in sync with each other.

But they have some sort of alliance that agreement that they form, because of which this whole ecosystems sort of works right. So, that is how I just try to club this 2 depending on the motivations that if two different kinds of motivations and of course, if educational institution can go on either of this, or can be both at the same time. So, it is not very water tight in that way. Does that answer your question? Ok and there can be many mores right

Student: It is a more like a shade of grey, right?

Exactly.

Student: (Refer Time: 20:45).

Exactly, yes.