Week 1: Introduction to Business Analytics

Video Transcripts

Video 1: Introduction to Business Analytics – A Motivating Example

Hello everybody. Namaste, Salaam and Sat Sri Akaal. Sudhir Voleti here and I welcome you to the first part of the first module, introduction and overview. I view Business Analytics as a vast and fascinating territory full of possibilities and opportunities. How better, in some sense to introduce you to this territory, than by walking you through a motivating example. Which is where I am headed next. Now this particular example takes me to Kenya. The venue is Kenya. The year is 2011 and the sector is dairy farming. What do you see on the screen? It's the picture of what exactly?

What you have there is a Kenyan tribesmen in the middle of nowhere. It looks like the hinterlAnd grazing cows and so on. There's no city. There's no habitation anywhere close by. What is interesting is what he holds in his hand a mobile phone. This will turn out to be very critical to where I'm headed next. Now, the sector dairy farming in Kenya faced the problem of uncertain yields from cattle assets. Now, these yields were uncertain both in terms of milk as well as in terms of money.

There were three main reasons and these reasons are not new or unknown and they are pretty common. They're prevalent elsewhere as well. The first start is biological cycles. Cattle gestation periods, for instance, which affect milk yield. Two, cattle feed, (quantity and quality), diseases - all of these affect milk yield. Three, volatile market prices, which affect the money yield for a given quantity of milk. All these three things. There were further problems too, which were more specific to Kenya. Farmers are small, cattle farmers in this case, they're dispersed over vast areas. Markets are small and localised and all of this information asymmetry was there, all of this leading to volatility in market prices.

This is taken from Google Maps, a map of Kenya. You can see the road network in Kenya. Around the capital Nairobi, you have good connectivity but if you look at the north, the vast hinterland out there is sparsely connected and a lot of cattle farmers actually, are in those areas. Now, silver lining to the whole thing, mobile penetration is high. In fact, way before India did, Kenya got into M-pesa, which enabled mobile to mobile money transfer. Now, into this scenario, entered Su Kahumbu. Stephanie Su Kahumbu, she starts a subscription information service called iCow.

That's the name of the service that she started up. Who is Su Kahumbu? Well, she is the daughter of a cattle farmer, of a dairy farmer and she's the fourth of seven Children. Grew up on a cattle farm. The first graduate in her family. Goes on to Oxford to do her post-graduation. Could have probably chosen to live on in England and make a good living there but instead she chose to return to Kenya because she could make more of a difference in Kenya. All right, now let's see what iCow was all about.

The iCow story and how analytics, Business Analytics as we will see, will play a part here. iCow started off with a very simple premise. SMS me information on these three issues. those three issues I had mentioned earlier in a standardised format. SMS not WhatsApp. She doesn't- This is 2011 firstly, and she is not relying on dairy farmers having access to smartphones.

Even feature phones will do. SMS me information on these three issues. in return, I will SMS you instructions on how to maximise milk yield. That was the basic, simple promise that iCow made. In the beginning, it was just information service. This is what it looks like today, registration for a cow on iCow, for instance, but back then it was very simple, the format and it was all standardised. How did it go? Well, word spread. This was the first year 2011 and in the first year, she had 42,000 people sign up. Any one person signs up in a village, the entire village tunes in. "Hey, how did it go? Did it work?" Think of the data asset that Kahumbu is building. She's basically, this is livestock data, health and yield data repository, quantity and quality data on dairy yields, milk production nationwide.

The kind of data that even the government would not have. Well, we will see. This is from their website later on, and you can see they've now branched into a lot of value-added services. But in the beginning, it started as purely an information service. Which brings up the question so, how does Kahumbu do this? You send me information on these three issues, and I will send you instructions on what to do to maximise milk yield. Right? The quantity of a feed, for instance, timing of feed and so on. How is she able to do this? Let's have a look at how this works, and this is where analytics enters the picture. In the beginning, she starts with little or no data.

She relies primarily on theory and guesswork. Later, when the data flow in, we will see analytics kick in. So in the beginning, her database, so to say, had really very little in terms of records. She's going to make predictions on what should you do to maximise dairy yield based on very little data in the first cycle. If her predictions actually improve yield, they don't have to maximise yield in the first cycle. They just have to do better than what was before. If yields have noticeably improved, what happens? She gets better customer traction if she gets better customer traction, more sign ups, more data, better measurement, better database records.

Finally, the database starts to get populated. The second time the cycle runs, she will have some data. She will know what worked in terms of predictions and what didn't in the last cycle. The third time the cycle runs, she will have more data and a much better idea of what worked and what didn't. Each time the cycle runs, the database grows and what she can do and cannot do, her prediction accuracy improves. By the time the cycle has entered the 5th, 10th, 12th, 15th, 20th generations, so to say, she is pretty much set. It's a virtuous cycle. After that, the model behind, the machine learning model that is basically, which is learning from the data and making the predictions has stabilised. It's become a pretty solid thing now and so on. Let's continue with the story.

Now, as a by-product, iCow became the most reliable database for non-farm businesses. So think about institutional and corporate dairy buyers. Say I'm an ice cream factory, I need to purchase milk of this grade and of this quantity, where do I go? Dairy markets in Kenya are small and localised. Turns out there is one information repository who would know, iCow. If I'm a veterinary doctor, so what happened was farmers started basically contacting Kahumbu and saying that, "Hey, my cow has fallen sick. Can you suggest, can you recommend a good doctor?"

So pretty much this became a platform that was connecting veterinary doctors with, sick animals and so on. Farm implements sellers and so on. So you had NGOs, the government itself, all of these. This became a platform, information platform for all these various stakeholders. Take a look at this, when drying off cow, give dry cow minerals- Unga brand or Vital by Bayer. She makes a brand recommendation, fortunes change in the market. Let's actually see how big a deal iCow is.

So could iCow organically expand? Well, yes, it could. into its subscription business into value added services. It became a B2B platform. This is later from their website and you can see all these different services that they are offering. What were the returns like at the users end? I can see what's in it for iCow, there getting information. They're building a data repo. What's in it for the farmer? Why would a farmer sign up? The average Kenyan farmer owns three cows.

Within seven months of using iCow, farmers report an average jump in yield of 33%, equivalent to owning a fourth cow within seven months. Which means, annualised, we're looking at a 50% jump minimum. There it goes. How about in terms of money? What was the service priced at? Turns out because there was a social service moto, she didn't price it very high. It was actually priced very low. In terms of money, for every dollar a farmer invests in iCow, the returns were 7,600% - $77. It doesn't make sense not to be on the iCow platform. Plus from the iCow side, there is a lot of head room to raise prices in.

The main point, all of this was enabled using just the humble feature phone at the users end, just a few laptops and computers at iCow's end. Of course, today it's a much bigger operation, and now they have servers and all that but before, that was all it took that was all that was needed. Low end analytics yielding huge returns and a bonus on the social side also a lot of benefits. I would encourage you to check out the iCow story. The possibilities this template brings up for other developing countries, including in India, and think about it, how many such domains are?

With just the mobile phone on one end, there is a lot of potential for information transfer, database building and that data feeding into a machine that learns from the data makes predictions, improves yield, the cycle continues. The format is doable now, even in open-source. Let me conclude the iCow example at this point. So Kahumbu, very first, the first point, she had a clear problem statement in mind. Optimise daily yield, which is your Y variable, the dependent variable, given inputs. So optimise the outcome - dairy yield given inputs which is those three issues that I'd mentioned. Now, to optimise the outcome she needs a model.

She needs to predict how inputs influence, inputs impact the outcome. She needs to predict what-ifs. What if this input falls? What if that input rises? What if both happen together? What if this one also changes? Each time you make a change. How does it affect the outcome? This process of predicting input changes to outcome changes, mapping them, that is modelling. She needs a model for her data. That model, the machine learns on its own. It doesn't require her intervention all the time. Data at the individual cow level well, basically, all of this is possible.

So it's a model that would make predictions for individual cows having learned from an aggregate of all the cows that are there in the database. These are algorithms that mine the data for patterns and then it's basically pattern matching. Very powerful as a framework, as simple as it is, powerful in some sense. Okay. The data that she uses is the primary asset, right? The raw material from which we basically extract a lot of insight. Here's the point, data alone is not enough. Analytics is also required, and I could be sitting on data and not knowing what to do with it.

This whole process of being able to connect that data, to impose structure on that data, to feed that data into a machine, to know what modelling is possible by the machine, to have some idea of how the machines learn, what it is they will learn, that is where we are headed, folks. The infra and the tech required for this operation is now fairly commonplace. She did this in the last decade, by the way

Today, it's much easier to do. Much faster also. A guy like me can basically code up what she did effective completely with open source materials, without having to break the bank. Are we ready to take these learnings from this example into the domain of Business Analytics? I hope you are because we are headed there. What data? What patterns? What problem statements, etc. might arise in this domain of Business Analytics? Let us have a look.

Video 2: What Is Business Analytics?

Alright, So what is Business Analytics? Now this question might seem terribly basic but please bear with me as I basically navigate an answer to this question that ensures both you and I are on the same page when it comes to certain fundamental terms that we typically use. Let's have a look at this. Now this is-- the curriculum is on Business Analytics and this raises certain conceptual questions, fundamental questions. What is a business anyway? I mean, when we keep throwing the word around, what does it mean? What is a business?

What are the objectives of a business? Which is a better way in some sense, to understand what it is a business does or tries to do or should try to do, in some sense. What is analytics? What does it do? What does it mean? What is data? Where is that coming from? Now this part will come as a by-product to where we are going, etc. So what I am going to do in the next few minutes is basically walk through these foundational, fundamental questions. Bear with me if they seem too simple, but let's go through this quickly anyway. Alright, the anatomy of a business, let's start with The Objectives of a Business, okay? Now I am taking the economic view of things.

There are two fundamental axioms in economics. What is an axiom? An axiom is a statement that can neither be proved nor disproved. It is what it is. If you choose to believe it, then what follows matters, okay? Every single theorem, every lemma in economics, when you boil it up, this is what is left, the axioms. There is no way to prove an axiom, it is what it is. In economics there are two axioms, this is one of them-- Firms exist to maximise profits. There is no proof for it, it is what it is. Economic as opposed to accounting profits but we will keep that aside for now.

Firms exist to maximise profits. In case you're wondering what the second axiom is, the second axiom says, 'Consumers exist to maximise utility'. Put consumers and firms together and what we have is a micro-model of an economy which we can now simulate effectively. All of economic starts from there, in some sense... yeah. Maximise profits. What is profit? You ask a high-schooler the question "What is profit?" and chances are the answer is going to be, 'Profit is equal to revenue minus cost', correct? Profit is equal to revenue minus cost.

Let us have a look at that equation. In fact, I want you to focus specifically on the cost side. What are the major cost heads in a business? What are the major costs in a business? Well you might say, know, operational costs, you have material costs, labour costs, regulatory and compliance costs, capital costs, working capital, and so on. All of that, all of these are costs. Now, take the OM domain, the Operations Management domain. The Operation Management domain... maintains and meets the fundamental axiom of maximising profits by minimising operational costs. That's basically how it contributes to the fundamental axiom.

Take the Corporate Finance domain, they contribute to the fundamental axiom of maximising profit by minimising costs of capital. Take the Accounting domain, they contribute to the fundamental axiom by minimising, of maximising profit by minimising compliance and regulatory cost, and so on and so forth. Let's call this the supply side of the profit equation. Why supply side? This is what will contribute to making the supply curve effectively.

Yeah, supply side. Now look at the revenue piece. The demand side of the profit equation is the domain of Sales and Marketing. How does Sales and Marketing contribute to the fundamental axiom? By maximising revenue and thereby maximising profit. So Sales and Marketing comes into this picture with a maximisation mindset, effectively. Now these different business functions Sales and Marketing, Operations, Accounting, Corporate Finance, All of these basically are one way to deconstruct an enterprise.

You take a business, break it down into parts, these are different functional parts of an enterprise. This is one way to look at what an enterprise is and what it does. It's not the only way, and there are other ways to look at it. For our purposes, this will serve. Where does market power derive from? Market power, in some sense, well I'll come to that part, but before I proceed, let me ask you this: These different functions that we saw, each of these yield their own analytics.

So you have Marketing analytics that basically helps the marketing function contribute to the fundamental axiom of maximising profit by maximising revenue. You have Operations analytics, which basically again helps the Operations domain contribute to maximising profit. You have People analytics, which basically helps HR and OB, Organisations Behaviour and the Human Resources function basically make people more productive and thereby lower overall costs and thereby contribute to the fundamental axiom. All of this coming together. Alright, which also brings up the other question of market power. Where does market power come from?

A firm typically will have market power if it specialises, if it does very well, if it is at the frontier of at least one of these two, typically one of these two. Either it is great on the demand side or it is great on the supply side. It is very difficult to be great at both without state support. Yeah. That is for another day. Let's proceed to the next big question now, okay? We are done with What is a business? What are the objectives of a business. Set it aside.

Now Let's come to "What is Analytics?" What is Analytics? If you web search for this definition, you will find a ton of definitions out there. What I am going to do now is for our purposes, define what is Analytics as part of Business Analytics. The anatomy of analytics. So anatomy, basically we're going to take this thing apart and put it together again.. Imagine you have a real-world system, the real-world system is nothing but our business, so to say. Your business. It exists in the real world. It faces real constraints. It faces real challenges. And you have real questions about what to do next, so to say. And then you have on the other side, real-world answers, real-world conclusions, answers to the questions you have.

Now I have questions, and there are the answers. In an ideal world, it would be a straight line connecting me from questions to answers. In the real world however, things are seldom that simple. In the real world, in order to get from real-world problems to real- world solutions, I take a detour. And as part of that detour Analytics comes in. How does this work? Well... I'm going to build a mathematical system which parallels the real system, okay? So I am going to build a model of the problem that I face, mathematical system, okay? Which parallels what I have in the real world. Now, abstraction is required. I can't model everything. I'm going to make a map, the map is not the territory.

I'm going to take important pieces of the business and build a model about how they interact... okay? Why is building a model necessary? Because once I have built a model, once I have built the mathematical system, yeah, I can do what-ifs, I can play around, it's like a lab now. I can change this input I can do that there, I can... and once I move things around, once I tinker things, I start to optimise the outputs. In other words, in my mathematical system, I can make a lot of changes and those changes will give me mathematical answers. Predictions of what happens, what will happen. Those mathematical answers, yeah, this process of getting mathematical answers to mathematical questions from a model to a prediction, that process is Analytics.

That process is Analytics. From this mathematical answer, now I have to translate it, interpret it, translate into a real-world answer now...okay? Interpretation becomes critical, in some sense. So this particular detour involves Abstraction, Analytics, Interpretation. Re-interpretation or translation, all of it. Basically the analytics' detour, so to say. It de-risks my job to have analytics. I can make a mistake in the mathematical world, no problem. To make a mistake in the real-world incurs real business costs. It makes sense to build a model... Okay? It may not always be possible to tinker and try things out in the real world.

There is a way to do that, though. And we will do that tool also as part of this curriculum. This last tool that directly takes you from real questions to real answers, okay? Which doesn't involve modelling, that one is called Experimentation. Experimentation is different from analytics. But we will do causal, my colleague Manish will cover this. Yeah, there it is.

What you see here now is the Analytics Course Map. Very broadly, so you have data. And you have data-driven modelling and data-driven learning, Data-driven Machine Learning, all of it coming in, broadly classified into two different streams: Supervised versus Unsupervised. Supervised implies that there is an outcome that is supervising the flow of inputs. Unsupervised implies exploration, effectively there is no one outcome that I can look at. All the inputs are also outcomes in their own right. In some sense, what I am looking at is exploration and structure.

Now the way this will work is we will cover some tools on the Supervised learning side and we will cover some tools on the Unsupervised learning side. We will cover both of them. On the supervised learning side, the very last tool that you see, Causal inference, is Experimentation. The direct route from real-world questions to real-world answers. Everything else is part of that detour, the analytical detour.

Video 3: Problem Formulation for Data Analytics

Let's now dive into problem formulation for Data Analytics. Problem formulation is a very crucial aspect of a managerial job. Problem formulation is so fundamental that everything that follows is impacted by it. When we formulate a problem, we are effectively building the box within which we will search for solutions. Build the box too small, and chances are that we may not find optimal solutions. Build the box too large and the effort and time required to search through the entire solution space may be prohibitively high.

There is an optimal. Let's get there. Let me start by showing you the anatomy of a business problem, okay, in some sense. Consider XYZ Corporation and these are the musings of a decision maker and country head for XYZ Corporation. One fine day, end of a long day. he or she is basically thinking these thoughts. Sales fell short last year, but sales would've approached the target except for six territories in two regions where results were poor. Of course, we implemented a price increase across-the-board, and so, our profit margin goals were just about met, even though sales revenue fell short. Two of our competitors saw above-trend sales. Still another competitor seems to be struggling, and word on the street is that they have been slashing prices to close deals.

Of course, the company, the economy was pretty uneven across geographies and those two regions in question, were weak anyway, particularly weak last year. And then there was that mess with that salesforce compensation policy that came into effect last year. One of the two weak regions actually saw a lot of salesforce turnover...

Folks, what do you think? What do you think? These are not exceptional things. These are everyday events in a manager's life, in a business leader's life. Yeah, and these are commonplace problems that basically, managers wrestle with. So, let me ask you a few questions about what we just read. 1. Is reality orderly? The answer is a strict no. Reality is messy. It is not orderly. Importantly, we have no right to expect it to be orderly. And unlike textbook examples, life seldom simplifies issues. Think about what happens in a textbook example.

A question is given to you in a textbook. The moment I see a textbook question I know that the information needed to solve the problem is there in the question itself. In fact, I also know which book and which chapter in that book it's coming from. So I know which way to start thinking and searching for answers from. In life though, things are seldom that simple. Where do I even begin? I mean, I don't even know.

Forget chapter, which book it is from. Yeah, these are the kind of real-life business problems that one would be called upon to solve. The medical analogy. Symptoms versus ailments. What is a symptom, folks? A symptom is the visible manifestation of an underlying problem, which is called the ailment. Think about the most common reason people go to doctors, let's say primary health care physicians. I mean, the pre-COVID data let's say.

What is the most common reason people go to doctors with the most common symptoms? Cold, cough, fever, effectively. Cold, cough, fever are symptoms. They are not the ailment itself. A lot of the times the body's immunity will fight the ailment, and what the doctor will do is give medicines to ameliorate the symptoms. If a business is sick, fighting the symptoms may not be enough to cure the ailment. We will see. The most common symptom in businesses, sales are falling. It's as common as cold, cough, fever. However, what is the underlying ailment we will have to diagnose it? We are headed there.

Okay. How do symptoms relate to ailments in the business example that we just saw? You have one symptom, sales are falling. And you have multiple possible ailments, multiple possible aliments. Oh, was it the salesforce compensation policy that went wrong or was it the economy that was weak, there's nothing we can do about that? Was that the problem? Was it competitor action? They were slashing prices to close deals. What was it? Was it our own product line, our pricing? What was the problem? We don't know.

Symptom is clear. Symptom is a visible manifestation of the problem. Sales are falling. What do we do? How to diagnose the problem in such cases? So, diagnosis is basically one looks at the symptoms. The doctor looks at the symptoms and tries to find out what the ailment is. In this case, how will we diagnose the problem? What is the problem? There are multiple possible causes and I think they're independent. They are actually interacting.

It could be some combination of those things. Things get more and more complex, don't they? When you go to a doctor, let's say with cold, cough, fever, what is the first thing a doctor does? What is the first thing the doctor does? They have this thing around their necks, right? The stethoscope. The first thing they do is basically check for chest congestion. If your chest is congested and you have cold, cough, fever, it emplies oh, the origin is bacterial.

Chances are, it's not 100%, there is no guarantee. Chances are it's bacterial. If the chest is clear but you still have cold, cough, fever, it might imply the thing is viral perhaps. Yep, again, no guarantees, but what it does is it moves probabilities towards one set of ailments rather than the other. This is important. The stethoscope is ready. It is available. It is cheap.

Fast, cheap, easy to use, low cost. Then and there, it is helping you to shift the probabilities of where the ailment may lie. In other words, in businesses too, we will try to narrow the field of ailments by ruling out low hanging fruit. Fast and cheap is ideal. Can data help? Yes, it can. We are headed there. Yeah, what data might be needed for the problem?

This depends on the problem as we formulate it. This depends on the box we will build. That is where I'm headed right away. We know that reality is messy, so there's messy reality right up there. And there's the analytics toolbox, the one, that stethoscope that will help me diagnose that will basically tell me what's going on in some sense. Now, in an ideal world, the way to go from messy reality to toolbox would be a straight line. Well, it's not an ideal world. In a real world, we'll have to take a detour.

Let's have a look at that. Consider a list of probable causes or product line is obsolete. Our customer connect is an effective, our product pricing is uncompetitive And there are more to sales for compensation and all of that, let's consider these three for now. It is core managerial responsibility to shortlist the major probable causes. How do you do that? Typically, you would use something called exploratory research. We will look at some exploratory tools that will help us extract these major problem heads these major potential ailments from that messy reality.

When we do that, we move from messy reality to something called a decision problem. A decision problem is a smaller, simplified version of potential version of what the problem could be. So DP number 1 could be, oh, should new products be introduced. DP number 2 could be, should advertising campaign be changed Customer connect ineffective, should advertising campaign be changed? Product pricing is uncompetitive, should product prices be changed? The moment you move from messy reality to a set of DPs.

Look at what we have done. We have imposed preliminary structure on messy reality. We have identified potential boxes within which we will search for solutions. DPs or Decision Problems may not contain sufficient information to map directly onto tools which are in the toolbox. We want to get to tools. So, another level of refinement may sometimes be needed. This level is what we call the RO, the research objective. A single DP can be broken down into multiple ROs, and each RO will map onto a tool.

From RO to tools is straightforward. Just to summarise what we did in this slide, there are two-way stations. When you want to move from messy reality to analytic tools, there is a detour. There are two stops to make. DP and RO. What we are doing is constructing the basic problem formulation framework. Let us look at the anatomy of a RO. Remember what anatomy is basically you're going to take this thing together and reassemble it. So you have messy reality, you have analytics tool box, and you have these two way stations in there. Decision problem, research objective.

A research objective, remember, because it maps directly onto tools, it has to be very precise. We cannot afford ambiguity there. A Research objective typically would have three components. Two components and a third restriction. It has something called an action verb. It has an actionable object, and it always fits into one line effectively. That third part is to enforce percision. Here is one example. Identify the real and perceived gaps in the product line vis-a-vis, our main competitors. Folks, what is the action verb here? Identify. What is the actionable object? Real, and perceived gaps.

Once we are here, the problem is well enough defined for me to open the analytics toolbox effectively. The challenge is getting to this place. Now, you might look and say, "But isn't this too constraining?" I'm in the problems that I face in my business, in my sector, in my vertical and in my domain are so unique. I'm sure you won't find an RO, an actionable verb, actionable object to fit them. Not true. Here is an incomplete list of verbs, and you could have more, right? Identify, define, describe, explore, generate, evaluate, select, test, you name it. And here is an incomplete list of actionable objects in any verb could typically connect with any object. The number of combinations, huge.

Chances are, any business problem you are facing, if well defined, will fall somewhere here. Once you define this, identify problems. identify opportunities, define the concept, generate hypothesis, evaluate potential test preferences, prioritise opportunities. The moment we come to this place, the toolbox is now ready to open and deploy. These complex, actionable objects are what we call constructs. Sometimes in academies. Identifying them, measuring them, modeling them becomes important.

And we will see a little bit of that along with the tools as we go forward. Let me take a deep breath at this point and quickly recap the story this far. We started with a motivating example, iCow. We saw how powerful analytics as a concept can be in a real business scenario, even in a low tech environment, in a developing country, in a sector that is thousands of years old, dairy farming. We looked at conceptual preliminaries. We went into what is the anatomy of a business? What are the objectives of a business?

We went into what is the anatomy of analytics? We looked at problem formulation for Data Analytics. We basically described a very simple problem formulation framework. It is important to be able to precisely formulate problems because then, the tools that we have will have maximum ROI, Return on Investment. If our problem is ill defined, then it would be pointless to blame the tool. We are going to head next into the Basic Data Analytics toolbox. And my colleague professor Ganguar will take over from here.

Video 4: Evolution of Machines

Professor Sudhir talked about the iCow example, in which how the data is used, how the technology is used to solve some of the challenging problems in the world. He also talked about the framework, how the business problems can be taken, and made into a Data Analytics problem. So, let's zoom out a little bit and see, think about how we reach where we are today. So, I'm going to start with this quote, "Our technology, our machines are the part of the humanity.

We created them to extend ourselves, and that is what is unique about the human beings." This is the code by Ray Kurzwell, he wrote, many books, but one of the very interesting book I would encourage you if you get time to read is, the singularity is near, talks about how the AI is a evolved. So as he says, the technology and the machines were always the part of ourselves, and we'd use them to extend our capabilities. So, if you look at how our tools have evolved over the time, so earlier we used to use the hand-hold tools, then we created the better tools, Lathe machines, bigger machines to do things faster, and now we have this automated tools, which can actually do things on their own pretty much.

Our sensors have evolved over the time, our communication devices have evolved. Right, instead of sending a data in the old time like zero, one bytes, now we can transfer the data through the undersea cables, tetrabytes of data, we can actually we went out in the space to collect data. As we see our computational powers have evolved, right? So, instead of The first computer, if you see was as big as the room. Those computers have shrunk into a small microchips.

Their power of computation has increased. Now, we are talking about the quantum computer, our ability to store information, now we have cloud servers, and the cloud storages where we can store zeta bytes of data. And in fact, if you think about the data, there is a lot of data which is generated right, there is a lot of data in the libraries, in the books, a lot of data is generated when we do transactions, the grocery stores, data is generated when we do the credit card transactions, stock exchange data, data coming out from the satellite, data which is created through blogs, Facebook, Amazon, Netflix, right.

There's a lot of data which is on the Internet, we actually are creating roughly 2.5 quintillion bytes of data every day, and we haven't talked about the IoT data. So, with all this data, still, we have so much of data still, we have to figure out to get the insights out of this data, and that's where the data needs to be organised, so that we can get some information, we can find some patterns in the data, and we can get some insights out of the data.

Video 5: What Is Artificial Intelligence?

So, there is so much of data we are generating. The question is how to analyse this data and that's where the idea of the artificial intelligence comes. So, what is artificial intelligence? Artificial intelligence refers to the simulation of human intelligence in machines that are programmed to think like humans and mimic their action. So, artificial intelligence is the goal there is to mimic the human intelligence. They can do things what, otherwise, humans can do.

So, if you look at this bunch of letters here and if you see these letters, what you will see is that you can very easily identify the letter. It is M, or O, or P, or Q. The question is, can machines also identify these letters? Right.

So, in other words, your ability to predict or assign a label to a new observation based on your past learning is what, is the generalisation is what we can think about is our intelligence. Or, one can think about as an intelligence, as for the machine purposes, when the transactions are happening, can you identify these are the fraudulent transactions and these are the right transactions? Can you predict who are my customers who are going to churn? Can you predict who will respond to these coupons and who will not? Or, can you look at some of these MRI scans and identify, is there a medical condition there? Or, you can think about another kind of intelligence. If you look at this picture, there's a picture of some pixels organised in a way.

Can you identify a dog in that picture, right. Or, if you look at on the right side, there are two words written, The and Cat. If you look at the middle letter, they are actually the same. But given the context, you can very easily see that in the first word, that symbol basically signifies the H, and in the second word, it signifies the A. So, this ability to make sense of an individual component in the context is what we can think about this is an intelligence. In the machine learning jargon, if you look at, if machine looks at the picture in the computer vision, can he identify this is the truck, this is the bus, this is the human, this the cycle, this the dog, right. Or, when you're looking at the some text.

Can machines identify these are certain kinds of objects there, these are the name of the cities, these are the name of the other kinds of objects, this is noun, this is pronoun, what is the relationship? And yes, now machines can actually do both kind of things. We have developed the algorithms in which machines can actually look at the pictures and identify objects in the picture. Similarly, in the natural language processing the machines can think of as reading the sentence. And they can actually not only read the sentence, they can identify different objects in the sentence, they can get the sense of what the sentence overall meaning is, is it a the positive sentiment, or the negative sentiment, and so on and so forth. So, the question is, when should we call a machine an intelligent machine?

And that's where the English mathematician who came up with this idea that how to define the intelligence. And he said, "A computer would deserve to be called intelligent if it could deceive the human into believing that it is human." Right? So, Alan Turing is supposed to be fore father of the AI and the computer science field. And that is the test, which is still annually perform on different kinds of task to identify how intelligent the machines are.

In this task, what happens is that there is a person sitting on one side of the table, and then on the other side there is, on one room, there is a computer, in other room, there is a person. The person C asks the question and then, machine responds to that question, and human also responds to the question. And if the person C can identify that, this is the human response and this is the machine response, then fails the test. But, if machine actually gives you the answer and the person C cannot identify or differentiate between these two answers, whether the answer is by the machine or by the human, then it's an intelligent machine. This is called the Turing test quite widely used even today.

So, if you think about it, with all this intelligence growing, machines are becoming smarter, right? The machines were able to, the IBM's deep blue algorithm in 1997 was able to beat the best chess player in the world. In 2011, IBM Watson, again, was able to beat the Jeopardy, the humans in the Jeopardy game. By 2016, the Google was able to create Algorithm where it was able to beat the humans in the very complex game called Go. And as we see as things have progressed quite a bit, right. Recently, AI created a portrait which was sold for $432,000.

Now, in 2018, Google came up with this new natural language processing algorithm, which can actually translate. So, if you say something to in English, it can translate that into the French. And that requires, I think about the different languages have different grammars. It's not about simply picking the word and converting that into different language. You have to reorganise the structure also, right. And recently, in Detroit Waymo, robot taxi made 1,000 trips, and these taxis are all AI-driven, no humans.

So, as we can see, the machines are becoming intelligent. And behind that, all this is basically the machine learning algorithms or AI algorithms. So, this is a very interesting thing, where Hans Moravec predicted that by 2050, robots will be able to actually execute 100 trillion instructions per second, and will be able to rival the human intelligence. So, far, we have seen that the machines are able to beat us in chess, they can self drive, they can do, they can start creating portraits, music, and stuff like that, and these things will keep on evolving. And at some point of time, the AI will be able to do things what humans can do.

Video 6: Machine Learning Techniques

So, let's look at the big picture overview of what is artificial intelligence, what is machine learning, what is deep learning? Remember, what is artificial intelligence? Artificial intelligence is machine's ability to mimic human intelligence, right? To do things which humans can do. This is nothing new. This all started 70 years back.

In 1950s also, we had machines who can actually make the decisions which were like the human decisions. But at that time, all these machines, the artificial intelligent machines there we used to call an expert systems. They were hard coded. So, you give them the set of instructions and you give them input. They follow through that instructions and give you the output. But the outputs, the input which is converted into output, almost like you know, that some human is doing that. And around 1980s, the whole paradigm shift happened. So instead of this, the new were using the statistical models, what researchers have able to do that that just by giving machines a data and the set of outcomes, machines were able to actually figure out what were the rules which are creating or converting this input into the output.

So, the techniques that give computers the ability to learn without being explicitly programmed to do so is what is the area of the machine learning is, and that is the sub field of the artificial intelligence. The sub field within the machine learning is the deep learning models. So, how deep learning models are different from the machine learning models? Around 2010, when the computational power reached to the critical level, we started seeing that there are certain problems. Within the machine learning, what happens is the output usually, and the input which is given to the machine is usually in the matrix form. It's There's a much structure data where we have columns and rows.

Rows are usually the observations, and the columns are the different features of the object. And machines basically read these features and convert them into the outputs. But humans need to still create and give machines that feature data. But there are certain kinds of problems which were not amenable for these kind of things.

For example, if you look at the picture. Pictures has a set of pixels. A 1MB picture will have so many pixels, and within that pixel, you can think about what are the things? Intensity, 0-1 variable, whether it's a black or white, what is the intensity of? If it's a coloured picture, you have to give them a RGB and this entire thing and just imagine one million pixels. It's a lot of data, and we don't even know because these data points are connected to each other. We don't even know how to extract the features out of these things. For those kind of analysis, the deep learning models came in.

The deep learning models do not require that we need to give them a feature data. It can take the raw data and convert that into the outputs and learn basically, machines are learning how this input data is converted into the output. The only thing in the deep learning models, because we are not guiding them as such, like in the machine learning model, they learn on its own. But the downside is that they require a lot and lot of data to learn these things. But now we have the computational power and the storage capacity to do that, and that's why you see that around 2010, that major shift happened, and that's where the big idea of the artificial intelligence gained quite a popularity among businesses and in the general public.

So, what we're going to do is we're going to look at the machine learning domain and how the machine learning domain is further divided. So, you can think of machine learning has three major parts. So, there are three different kinds of machines learning.

One is called the supervised learning. The other one is called the unsupervised learning. Then there is a reinforcement learning. So, what's the difference between supervised learning, unsupervised learning and reinforcement learning? Supervised learning, you can think about is used for the prediction. So, there is some kind of outcome you want to predict, and the outcomes could be a continuous variable.

For example, I want to predict what is the temperature going to be tomorrow and the temperature can take any value. Temperature could be 28 Celsius, or 30 Celsius or 28.9 Celcius, so it's a continuous variable. In those kind of settings, we use the regression. Or you can think about, the outcome could be ordinal variable or could be a categorical variable.

For example, if the customer is coming to my store out of these four brands, A, B, C, D, which brand he will buy? So, that's the outcome I want to predict, which can only take four values. That's the classification problem. So, that's where the prediction is happening as a categorical variable. You can think of I show the machine a picture and machine needs to identify, is it a cat or is it a dog? That's a classification problem. So, that all falls into the supervised learning.

What is unsupervised learning? Unsupervised learning is another domain where we don't have the output per se. What we want to understand is the latent structure in the data and the two primary kind of examples there are. Sometimes, we want to understand the structure so that we can actually compress the data. We can reduce the dimensionality of the data. So, think about the picture. When you click the picture in the raw format, it's a very big file. It's a lot of information. Can I actually compress that file? Can I reduce it to lower format so that still I have all the information intact there, and I have removed unnecessary noise?

So, that's an idea in the dimensionality reduction. Now think about, I have a data which is multi-dimensional, I have 100 dimensions, and I want to visualise that data. How are you going to visualise? We cannot visualise things in the 100 dimensions. We can only visualise things in the 2 dimensions or 3 dimensions at the best. So, how to basically take those 100 dimensions and create 2 dimensions which keep most of the information so that we can understAnd what does the data looks like? That's the dimensionality reduction idea.

So, other kind of unsupervised learning is the clustering. What is clustering? When we want to group things together. So, think about, the way we understand the world is things are grouped. This is how we will look at the world, that these are the trees, these are the cars these are the humans, these are the animals.

Now, here we already have figured out how to organise things. But think about the web pages on the internet. The web pages on the internet could be of all different things. But keeping that data and trying to understand and get insights from the raw data will be very difficult. So, you want to organise all this data into some structure so that we can understand what is the pattern there. Maybe we can get some insights by doing that. For example, we have the bunch of customers. We want to segment these customers into smaller groups so that we can serve them better. So, this is all the idea of the clustering analysis.

The third is about the reinforcement learning. What is the reinforcement learning? Reinforcement learning idea is largely used in the robotics, and now getting used also in the things that you saw that, for example, the IBM was able to beat the chess player or the Google Go AI was able to beat human in the Go game is all in some senses the reinforcement learning. But the way to think about the reinforcement learning is like the kids when they are playing those maze puzzles.

In the reinforcement learning, the machines are only given the instruction that this is your goal, and these are the rules. So, you want to go from place A to place B, and you can take one step at a time, you can turn left or you can turn right, and there are all different possibilities there. Some of the possibilities will lead to the dead end, and some of the possibilities will lead to the outcome. So, the only thing given to the machine is these are the rules, that you can take left turn, or you can take right turn, or you can take one step forward or one step backwards, and the goal the machine is to achieve is to reach to some point. Then machines kind of figure this out on their own, and that is what comes under the reinforcement learning.

So in this course, we're going to largely focus on the supervised and the unsupervised learning models. So, one another way to look at the supervised learning models and the unsupervised learning and the reinforcement learning model is, the supervised learning models requires some outcome variable, which we are interested in predicting often and to do that, what we really want to understand is the relationship between the input variables and how it is related to the outcome variable. So, the goal here is to discover that relationship. So, that is what the supervised learning is about.

What is unsupervised learning? Unsupervised learning is there is no explicitly given Y variable. We want to understand the latent structure of the data, and that's where if I can understand the latent structure of the data, we can reduce it bring it to the lower dimensions, compress the data. Compressed data can be sent through the internet much efficiently, or in other case, we can cluster them into different groups. the reinforcement learning, as such as I said, requires a lot of deep learning models.

But it's used in the self driving car, robotics inventory, for example, how to teach robots to do riveting. It may look like a very simple idea that how do robots actually are able to screw the nut? But it's actually when you think about it, it's a very very difficult problem, which otherwise are so easy for humans to learn for machines to figure this out. These things are gaining traction when we're thinking about the automation, and that's where the reinforcement learning ideas come very handy.

Now, again to get the big picture of you, there is a place where you can find a lot of information about, at least the big picture view about a different kind of machine learning models which are used in the business. So, there is a link on your LMS, which you can go. This is the executive guide to the ML which is created by the McKinsey. It talks about deep learning, machine learning within that supervised and unsupervised learning and different examples in the business context and different algorithms which are used to do that.

Video 7: What Is Supervised Learning?

Okay, so let's look at one of the examples of Supervised Learning. Here, we're going to look at the classification example. So, think about... the doctors need to identify somebody that whether you are obese, or you are normal or you are underweight, How do they do that? They basically measure your height and weight and then convert that information into a some kind of a BMI index. And using that BMI index, they can predict whether you are overweight or you are normal or you are underweight.

In this whole thing, what happened is how the input data, which is the height and weight, is converted into the obesity level, or this classification that you are obese or underweight based on some kind of formula. And where did this formula came from? This formula came from the theory, from the past knowledge, which doctors have gained, right? Now so, this is called the traditional knowledge-based classification. In this, what happens is that the theory is used and the input variables using that theory and the formula is converted into the output variable.

Now, how it is different from the machine learning? In the machine learning context, the things actually are flipped. In the machine learning context, the theory doesn't identify the rules. Rather, the machines are going to figure out the rule for you. In this case, what you will do is you will actually have the input data, height and weight of thousands and thousands of people and then the obesity levels. Right? So, you will give them the input and the output. And the machines will figure out this formula: how to convert this input variable of height and weight into the obesity level.

So, that's how the traditional knowledge-based classification is different from the machine learning classification. So, let's look at the other example of machine learning classification. So, suppose I give you these three pictures and I tell you, these are the picture of one kind of flower. There's a picture of another kind of flower, and there's a picture of another kind of flower.

These three flowers are of the similar kind, but they are kind of a little different. One is the Iris Setosa, Iris Virginica. And now the question is, if the expert looks at the picture, he can say, Okay, this is the Setosa flower. This is the Virginica flower. But can we teach machines to actually identify these things? And the answer is yes. And here again, there are two different approaches, which I'm going to talk about.

One is the simple machine learning approach. And then, I'm going to tell you how it is different from the deep learning approach in the same problem. In the machine learning approach, what we will do is we kind of, you know, guide machine. In other words, we tell that the machines, you know what, if I give you the length of a petal, length of a sepal, width of a petal and width of a sepal, that somehow actually is good features or the variables, which can help you identify whether tis a Setosa or a Virginica flower.

So, in other words, what we will give to the machine is these four different features. Sepal length, sepal width, petal length and petal width. And then using that information, I also will tell machine that these are the flowers which are Setosa. These are the flowers which are Virginica, and with this whole data, machines will figure out the relationship how to convert these four features into the outcome Variable. So, this is a machine learning kind of you know, simple machine learning tool you can use.

Or in the same case, what I can do is I can ask machines to just directly, instead of giving them the feature, I feed the pictures of these flowers, right? And these pictures are all kind of, you know, marked that these are the pictures of the Setosa flower. These are the pictures of the Virginica flower, and these are the pictures of the other kind of flower, right? And then they analyse the pixel, right? The pictures are basically the bunch of pixels. And then, using that pixels directly figure out next time, if I show the machine the new picture, they will be able to classify whether it's a Setosa or Virginica.

The difference between the deep learning and machine learning model is in the machine learning model, we gave them features, which we thought were important, for example, sepal length and sepal width and petal length and petal width but did not give them the rule. The rule that they figured out from the features. And how? What kind of... how to assimilate all these information and give you the output. In the deep learning model, we didn't even give them the features. We give them the raw data, and the machine has to figure out how to convert the raw data into the outcome variable, which is the classification whether it is a Setosa or Virginica flower.

So, what machines think about that, this is-- I'm just plotting this kind of data in the two dimensions... bunch of flowers, the blue flowers, green flowers and the red flowers. And then you have the sepal length on the horizontal axis and the sepal width on the vertical axis. What machines have to figure out these boundaries on their own. Right? So, these boundaries machines have to figure out. This is how basically, the machine learning classification model can be very intuitively there. Think about... you can think about those as just identifying these boundary things, right?

Okay, Let's look at another example, which is more closer to the business context, right? So, think about the government wants to kind of scrutinize all the tax filings and then figure out these are the kind of tax filings, which are normally, we don't have to further process them. And there are certain kind of, you know, tax filings, which needs a further scrutiny. So, what they will do is they will have the expert who will look at all the taxes and then say, okay, these are the... This looks okay. And we don't need to further scrutinise them. These are the tax filings, which need further scrutiny. Right? So, experts can do that. How do they do that? Really, we can think about intuitively. They're looking at bunch of variables, a bunch of features in the tax filing to figure out, okay, these are the things, which are suspect.

And these are things, which are not suspect. Maybe they're making some many decisions, right? If they see the refund, whether they are the tax filer has asked for the refund or not. And that's a good variable to figure out whether this person is basically requires further scrutiny or not, whether they are cheating on the tax or not. Right? So, think about this problem. There are 100 million taxes filed every year, and we need lots and lots of experts to go through each and every tax filing and kind of, you know, classify them whether these filings require further scrutiny or not. Can we train machines to do that? And the answer is yes. How will we do that?

We will feed in the machine all the past from the last year when the experts actually made these calls, we can give them for each and every tax filing, we can give them bunch of information, right? Whether the features would be, for example, whether this particular tax filing ask for the refund? What is the marital status of this particular person or the tax filer? How much is the taxable income, right? Based on this information and we also tell them whether this particular tax was classified as further scrutiny, yes or not to be scrutinised, right? And then what we will do is we'll ask machines to create some kind of rule, which humans use. So, that is their approximation of what humans... how the humans made decisions, right?

They will create some kind of decision tree, some kind of if-and-else statement, if you want to think about, to figure out how each and every observation is classified, whether needs a further scrutiny or not, right. So, we can train the machines to do that. And we will learn this technique in one of the classes later, right? So, once we-- machines have figured out that this is the rule, then we can use the same rule for this year. We can classify each and every tax filing, we can feed in the data, and the machine will be able to predict whether this requires a further scrutiny or not.

So, you can imagine, instead of having thousands of experts going through each and every tax, one single machine can identify all these tax filing and classify them for further scrutiny required or not, right. And that reduces the kind of, you know, work for the humans. And that's how the machines are able to help humans improve their efficiency and do a job, which otherwise will take months can be done now in probably in days.

Video 8: What Is Unsupervised Learning?

To better understand the unsupervised learning, I'm going to walk you through one of the examples of clustering. So, think about all the data, all the information in the world is always organised in some structure, right?

For example, the life is divided into domains, which are subdivided into smaller groups. And that's how we can identify, this is bacteria, this is plant, this is human, this is fish, this is fungi, right? Why we want to do that? Because having the structure helps us, kind of you know, very quickly figure out something about the particular new observation that we have. This is about-- this is a fungi. We know quite a lot about what, how it will be, right. So, these clusters have certain properties, right. And that's how these clusters are created.

But all these clustering, how these domain and all these clusters have been created, is by some kind of, you know, a lot of scientists kind of, you know, did lot of research. It took them a lot of time to figure out that, how to classify, how best to classify the life on Earth, right? How should we, what are the things, which we should call bacteria, which are the things we should call fungi, which are the things we should call animals, and which are the things we should call plants, right. There's all based on the domain knowledge and the theory.

Now, think about, similarly, if you think about all the elements on the Earth are classified on the periodic table, right. And how they are classified? Based on their how many -- how many atoms they have, right. What, how many neutrons and protons within the atom they have, right. And once we have organised all the atoms into particular kind of in the groups and the structure, it's very easy to identify the properties of the -- if they identify the new mat-- new kind of substance on Earth, where you can very quickly say, okay, this is what the atomic structure of the substances is, so, this is how it will behave like. Whether it is, it falls into the alkali group, or is it falls into the mineral group, or the halogen, or is it the inert gases, things like that, right? Because different kind of, you know, groups behave very different. Some of these things are very volatile. Some of these things do not interact with the other metals and so on and so forth.

So, having a structure on the kind of, you know, data is very useful because it can give us a lot of insights and then, we can use that insights to kind of, you know, understand something more about the each and every element within the cluster. But the thing is, not everything we know is very well organised, right? And to organise all these things, it requires a lot of efforts in. So, can we ask machines to kind of, you know, do this job for us? And that's where the cluster analysis comes in.

For example, now we have so many web pages on the internet. And we want to organise them into some order so that we can retrieve them very easily, right. So, for example, Yahoo Directory has organised all the web pages in these different categories. How do we identify these categories, right? Now, think about our other examples. How many and what type of web pages are there on the web? How many types of customers are there in the market? How many types of people are on the social network and how do they behave? Because, how we going to cluster these people into the social network? Maybe, you know, one way to cluster them is that the people who do, kind of, you know, who are on the social network and post lot of things very often versus the people who do not interact on the social network that much, right.

Or, we want to, kind of, you know, look at the different genes, human genome, and classify humans into different kind of groups to identify these are the people who are more susceptible to certain kind of disease, right. And then, we can find, and if we have the new human, we can say, okay, you belong to this group. And that's why I would recommend you that you should take care of your health because you are more susceptible to diabetes or something like that, right.

Similarly, we want to organise all the different songs into different kinds of genres, different kind of movies, and that's what you see when you go to the Netflix. Just imagine if there was no kind of sorting there, right. How difficult it would be for you to find out the movie which you would be interested in, kind of, watching at that time. And that's what the clustering or organising thing into a different structure is very important.

The other reason why you want to cluster the things is because you want to, we want machines to actually-- so, think about, suppose I give you a bunch of basket of fruits. And say, you know, can you club them into the different groups? What you can do is you can actually club these groups into based on the colour. These are the green fruits, and these are the red fruits and these are the orange fruits. You can actually cluster them based on, these are the round kind of fruits and these are the long fruits, and or some other way, or the combination of these things.

So, within the 2-3 dimensions you can see, you can cluster things into very very different ways. Some of these different ways of clustering are very useful, and certain kind of clusterings are not useful. And that's where, unless either theory can guide us that may be all the elements should be clustered, clubbed based on the atomic structure, or in certain cases, we really don't know, as of yet, how to basically put different things into the clusters.

Now, one kind of clusters could be very useful or other kind of clustering could be not very useful, right. And that's why the clustering is some kind of, you know, one way to think about is, exploratory in nature, right? So, you give machines the bunch of, again, the data features. For example, in the fruits case, you can give them this size, weight of the fruit, the colour of the fruit, whether it's round or oval or long, and these are the different features of the fruit, and ask machines to club these particular set of fruits into the different bins, right. And they can come up with, maybe machine will, will club them based on the size, maybe we will club them based on the weight, or maybe some other combination. And then, you can look at these different clusters.

Certain clusters are not very useful, but some clusters may look out very useful. And then, you say, wow, I see a pattern here, which I never thought that there is a pattern like this, right. And you can use that insight to do something more, right. It will, it can be used for further, for the supervised learning purposes, and so on and so forth. So, in nutshell, what clustering is one kind of unsupervised learning, which is, basically, grouping similar things together, right. And why we want to do that? We want to understand or discover the latent structure, right. Maybe sometimes, we want to summarise the data.

Now, instead of having thousands and millions of observations, now, we basically say there are five different kinds of fruits. These kind of fruits have these properties. Most of them are heavy. They are round and they are yellow in colour, right. And these are the kind of fruits who are basically a green. They may be oval in shape, and they might be citrus in taste, right. And then, kind of, you know, it's easy for you to figure out okay, these kind of fruits might be good for certain kinds of things. These kind of fruit maybe are good for some other kind of stuff, right. So, then you can also, once you have clustered the things, you can put the labels on them.

Again, this is all about organising the data into the, in a way so that it's more informative for you, and for you to kind of get more insights out of the data. Similarly, you can think about instead of having one million observations, if you have only grouped this whole observations into 10 different groups, you can just look at the average feature for each and every cluster, and that becomes your representative group object, right. Now, you have to only think about only five objects instead of thinking about one million objects, right.

So, these are the different places where the clustering is used and can be used in the business context also. For example, I have a customer's database, right. The customers have so many different features, maybe their age, income, geographical location, all different kinds of things, their education level. I want to club them into such a way so that I can identify these are the people who are more likely to go for the kind of products I'm creating. These are the customers who are good customers for me. These are the customers I should not be worrying about, maybe because of they are not falling into the certain income group, and they cannot afford my product, and these are the people who are not interested in my products, so on and so forth, okay.

So, just to give you the big contrast between the unsupervised learning and the supervised learning. In the unsupervised learning, where the clustering, again, we use the features, right. We use some factors, which described the particular object, but we really don't know which class they belong to. We want machines to figure out, that group them in such a way that within the group these things are very similar, and across the groups, they are very different. Okay, this is the goal we gave to the machine, and then, machines organise all these observations into the groups, right. That's the clustering. Here, there is no outcome variable ex ante, right. Unlike that, in the classification, again, we give machines the bunch of features, the same features, right. But there, we know that this particular observation belonged to this class and this particular observation belonged to this class.

And the goal there is when the new observation comes, I want machines to basically predict whether they belong to class 1 or class 2, right? So, we looked at the both the examples, and I hope the distinction between the supervised and the unsupervised learning is a little bit more clear with this example.

Video 9: Introduction to Business Analytics: Summary

So, I hope by this time you have a good picture of what the AI is, what is Machine Learning, what is deep Learning, how to think about. We covered lot of examples, and throughout this course, we will talk about different Machine Learning techniques which will help you make better decisions. But largely just to think about how will you apply these Machine Learning and Deep Learning models in your context. So, the way to think about is, think about what decisions am I making more often?

So for example, you are making decisions about which offers to send to which customers. Then the question is, How you are making these decisions? What is the basis? How do you figure out that, are you sending the coupons to everyone? Or are you sending these offers to selected people? How you are deciding which are the people who should be sending this offers to and who are the customers you should not send the offer to?

So, you can maybe looking at the past purchase behaviour of this, how often they came to your store, how much money they spent? When is the last time they bought anything from you? And maybe these are the things which you are using to make those decisions. Those are called the basis for making the decision. And then, you also need to think about how you are measuring the success, whether whatever the campaign you did last time, are you measuring whether this campaign was successful? You may have spent lot of money in doing this campaign? Are you getting the returns or the money, right?

So maybe you are basically observing that okay, I sent out one million coupons and I saw that only 2,000 people got motivated and came to my store and kind of availed that coupon. So, what is the metric of success? So these are the kind of business, think about these are the-- from the business side, how you think about the decisions? Decisions, what is the basis of the decision? And how do you measure the success of that decision?

Now, when you convert that into the Data Analytics problem, the same thing you're going to do. What you're going to start with this, first thing is, What kind of data I need to collect to evaluate my decision? Again, what kind of data is the basis variable? That you basically were using the past purchase behaviour to figure out who to send the decisions and based on that decision, you basically measured how many people ended up availing the coupon. So this is how you were evaluating the decisions. How many customers redeemed your coupon?

Now, think about, okay, If I'm trying to use Machine Learning to improve my decision, I can actually feed in the machine all the information I used to make the decision or maybe there are certain things which I could also use to help my machines make the better decisions. What are those variables? What are those features? Do I have that feature? Do I collect those features or not? If not, how will you get that feature? You can use those features to improve the efficiency of the machine. And then of course, when you're thinking about the Data Analytics, the other thing you have to think about is what kind of model?

So of course, this particular problem is more like a supervised learning problem where based on the past purchases, you want to predict whether this particular customer, whether I should actually send him a coupon or not? But in other cases, this could be an unsupervised learning problem. So, what kind of supervised learning models I will use? Like, Should I use a Decision Tree? Should I use the Logit Model? Should I use Regression? That is the one thing which you have to figure out. We're going to cover all these different kinds of supervised techniques in the later modules.

So, I hope with all this, it gives you a little bit of sense of how the business decision problems, how you should think about the business decision problems, and how you should think about converting them into Data Analytics problem.