Introduction to Industry 4.0 and Industrial Internet of Things

Prof. Sudip Misra
Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur

Lecture – 37
IIoT Analytics and Data Management: Machine Learning and Data Science - Part 1

In the last lecture, in this module I had given an overview of analytics for IIoT and in this I am going to give you the overview of machine learning. So, we are going to talk about what machine learning is, what are the different types or methodologies in machine learning which are very popular and thereafter I am going to give you little bit of more idea about some of the different popular techniques that are there.

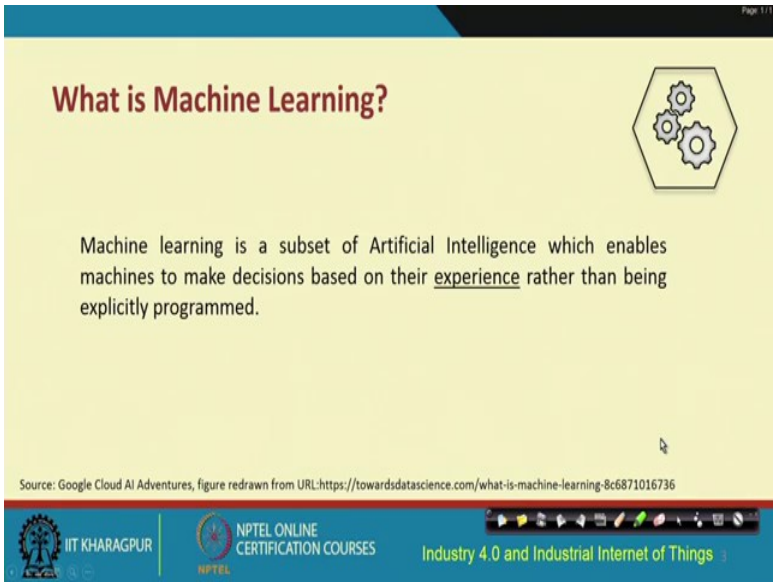
In this, I am going to give you only an expository view of machine learning. We are not going to go through any of these methods of machine learning in detail because the whole purpose of this course is just to expose you to what is out there with machine learning which can be applied for solving some of the problems in IIoT.

So, the purpose of this course is not to really get into the depth of machine learning and if anybody is interested and wants to have knowledge of machine learning and the different techniques that are there for machine learning. There are dedicated courses; courses in NPTEL which could be referred to for getting in-depth understanding of machine learning and the different methodologies for machine learning including deep learning, AI and so on.

So, in this course we will be at a very higher level and we will try to get just an idea about what is machine learning; what is what; what are the broad techniques that are there in machine learning and so on.

So, let us first try to get the ideas of the basics of machine learning.

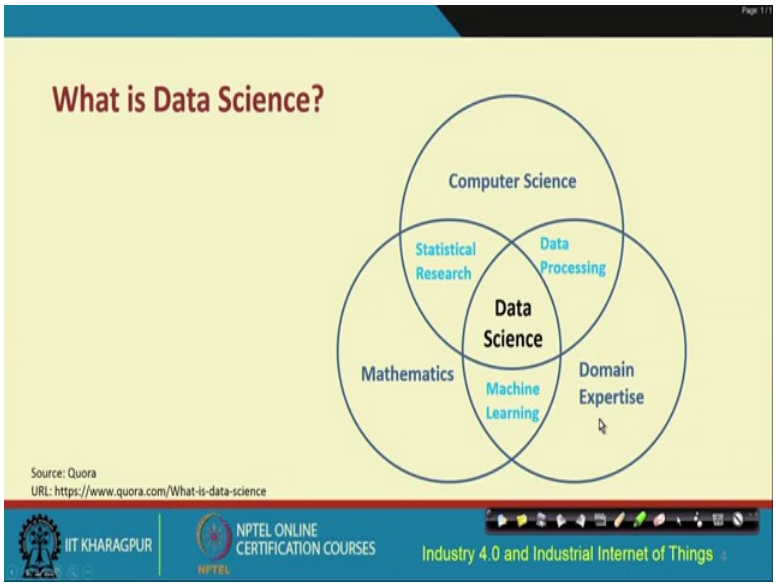
(Refer Slide Time: 02:07)



So, what is this machine learning all about? All of us we have heard about machine learning nowadays. Machine learning is very popular. Machine learning being used for analytics. Analytics very attractive for IIoT industrial, IoT applications; so, machine learning is nothing, but it is a subset of artificial intelligence. So, artificial intelligence is a branch of computer science which talks about how to basically imitate some of the natural intelligence that is there and create an artificial scenario of intelligence with the help of different computational devices with the help of different software and so on.

So, we have spoken about artificial intelligence briefly we have got an overview of artificial intelligence in an earlier lecture and so, machine learning is nothing, but it is a subset of artificial intelligence. So, machine learning talks about that how you can try to harness the experience from the past and try to make decisions for the future. So, machine learning basically will enable the machines to make different decisions based on the past experience rather than having the machine perform the actions based on what it is explicitly programmed to.

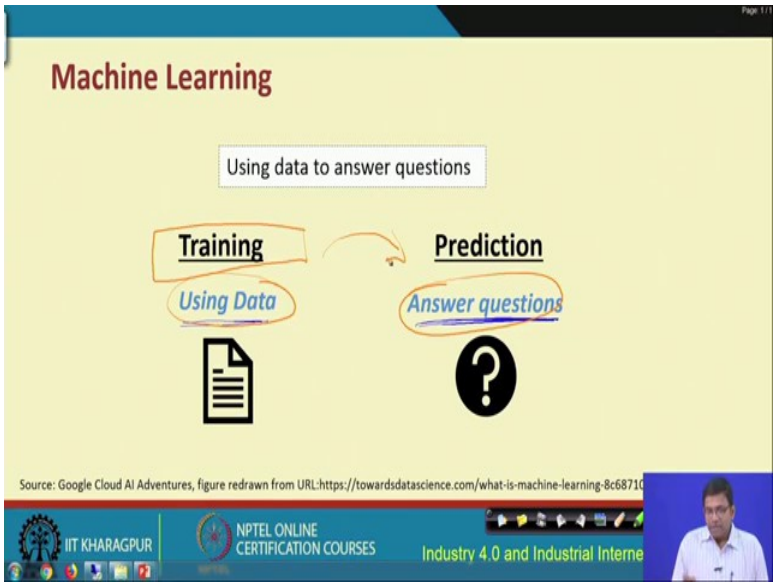
(Refer Slide Time: 03:43)



Data science; that is another term that is being used popularly at present. So, everybody is talking about machine learning, data science and so on. So, how do they position each other with respect to the overall knowledge? So, this is the overall scope of machine learning and data science with respect to branches such as mathematics, computer science and so on. So, basically data science combines the knowledge from mathematics, computer science and domain expertise. So, it is an overlap of each of these the domains. It uses techniques from each of these; it uses statistical techniques, data processing, machine learning techniques and so on.

So, we can think of data science to be at a higher level which is convergent, which is basically an intersection of the disciplines of mathematics, computer science and also the knowledge from the domain expertise. So, this is the overall positioning of machine learning and data science.

(Refer Slide Time: 04:57)



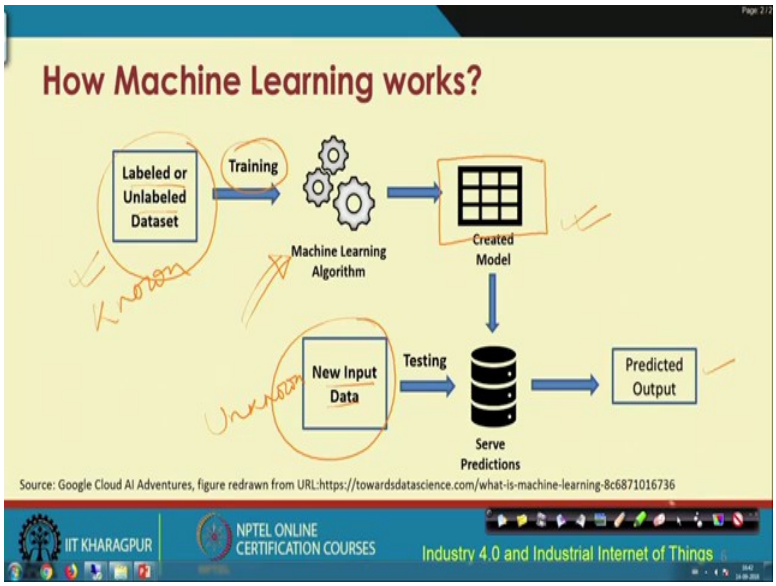
So, what is machine learning? As I told you earlier that we are talking about predicting something in the future based on the past history, past knowledge and so on. So, the whole idea is that you use the past data, existing data you use and then you try to make predictions or answer certain questions for the future, right. So, this is the whole idea behind machine learning.

So, in machine learning basically what you are trying to do is based on the past data you are trying to predict or answer questions for the future. So, what we have is that we need some kind of a training data set which will basically be the data of some observations that were taken from the past and based on that past experience try to predict the future this is what machine learning is all about.

So, this is one view of machine learning, a popular view of machine learning. There are different-different other views of machine learning as well, but making predictions in the future is something that is central to the theme of machine learning. How you make these predictions; what kind of data how much of data; whether data will at all be required or not. So, there are different-different questions people ask, people who are working on the different research themes of machine learning they try to address all these different types of varieties of questions.

So, let us move forward and try to understand a bit more about machine learning.

(Refer Slide Time: 06:55)

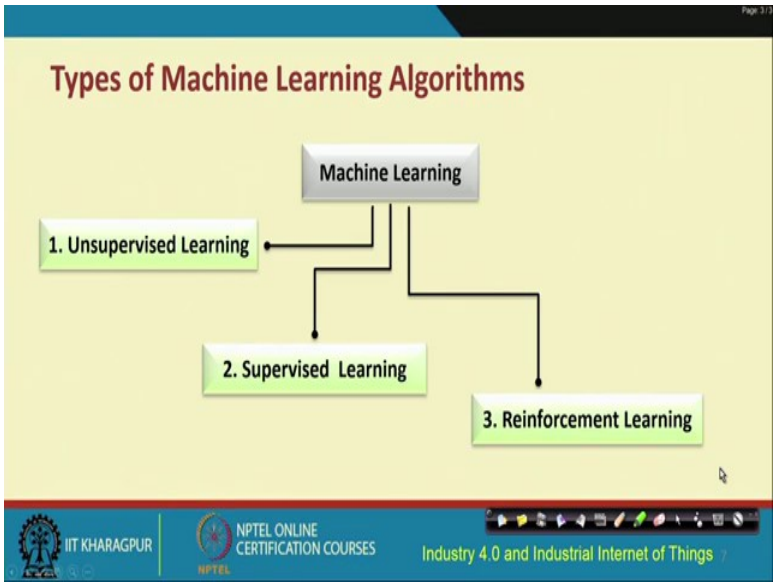


So, how does machine learning work? So, for machine learning what has to be done is that you need some kind of training data. You need some kind of training data this is basically the past historical data for instance which are labeled or unlabeled data and you design certain machine learning algorithms which will take this data, train the algorithms based on this data and will create certain models. And, will create certain models and those models are the important ones.

So, these were known data these models that have been trained and created will be used to predict something in the future for unknown data. This is the unknown data, this was known data. So, this created model that has been created based on the known data will be used for predictions for the new input data which is the unknown data.

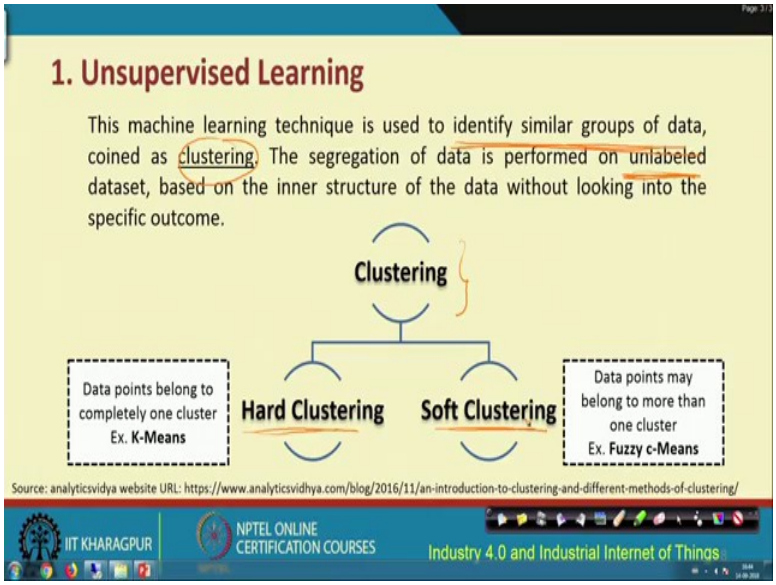
So, this is basically how this prediction is done in machine learning.

(Refer Slide Time: 08:31)



So, there are different machine learning algorithms they can be classified broadly into different types. These are the three main classifications that are very well known unsupervised, supervised and reinforcement learning which is kind of a semi-supervised kind of learning. So, we have unsupervised, supervised and reinforcement learning machine learning algorithms.

(Refer Slide Time: 08:57)



So, unsupervised learning; in this the machine learning happens in this way. So, there are similar groups of data which will have to be classified. So, in unsupervised machine

learning what you try to do is you try to identify similar groups of data and this process is known as clustering. This is one popular unsupervised learning; clustering is a popular unsupervised learning technique and this basically will help you to classify the data into similar groups.

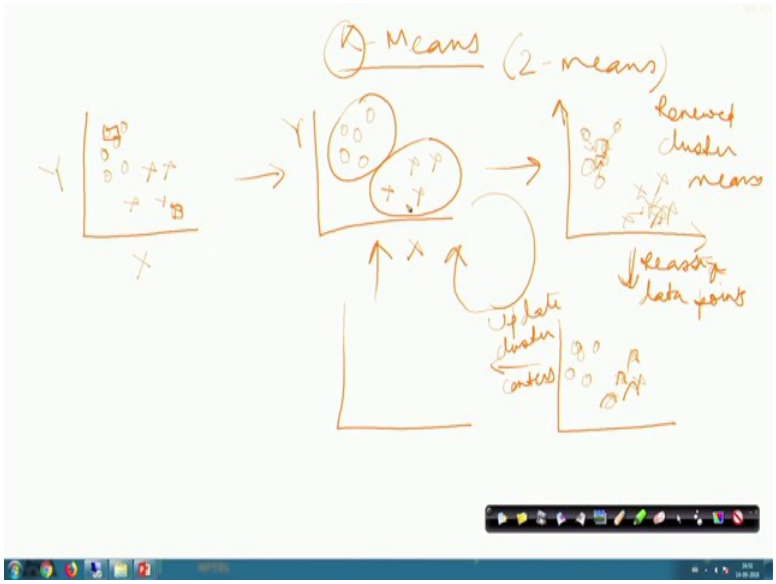
So, this classification or segregation of the data is performed on unlabeled data set based on the inner structure of the data and here basically this is very important unlabeled data set, right. So, this is unlabeled data set on which this unsupervised learning works and that is based on the inner structure of the data without looking into the specific outcome.

So, there are two main classifications of clustering algorithms one is known as hard clustering and the other one is known as soft clustering. Hard clustering basically what it does is it clusters into different distinct groups. Soft clustering on the other hand may have overlaps of clusters. Certain points may belong to two or more clusters together whereas, that cannot happen in hard clustering.

Algorithms; popular algorithms such as the K-means is a hard clustering algorithm, where the data points will belong to one cluster completely or another. In soft clustering, soft computing techniques such as fuzzy logic are used in order to come up with faster classifications, faster clustering, extremely faster clustering, but the classification is not hard, and techniques such as fuzzy c-means which is based on the concept of fuzzy logic is an example of soft clustering where the data points may belong to multiple clusters, right. So, these are the two main techniques.

Now, let us talk about this K-means. You might be wondering that what is this K-means all about. You know even though I do not intend to give you a definitive idea of all these different machine learning algorithms in this half an hour course, but I think in order to keep things in perspective and to motivate you enough one of these basic algorithms the K-means clustering algorithms I can give you little bit of idea about how it works. So, this is how this K-means algorithm works.

(Refer Slide Time: 12:17)



So, K-means basically the way is that you want to let us say that you have certain points and you want to classify these different points. So and let us say that this is your X and this is your Y. So, K refers to any integer.

So, let us say that we will talk about the 2-means algorithm. So, 2-means would mean that we are talking about two clusters, two centroids. So, we will start with two centroids which will be initially like the anchors to start with. These are kind of anchors to start with. And, so, you start with this you randomly select these anchors, the points and then you proceed further. So, then what we do we take each of these points 1 2 3 4 etc. 5 6 whatever and we measure the proximity of each of these points to these anchors.

These anchors are like the seeds. So, we start with these seeds or the anchors, we try to measure the proximity of each of these points in this space in this 2D space to each of these anchors. So, let us say that finally, we get something like these circles are the ones which are closer to this particular anchor whereas, these cross marks are the ones which is closer to this particular anchor or the seed.

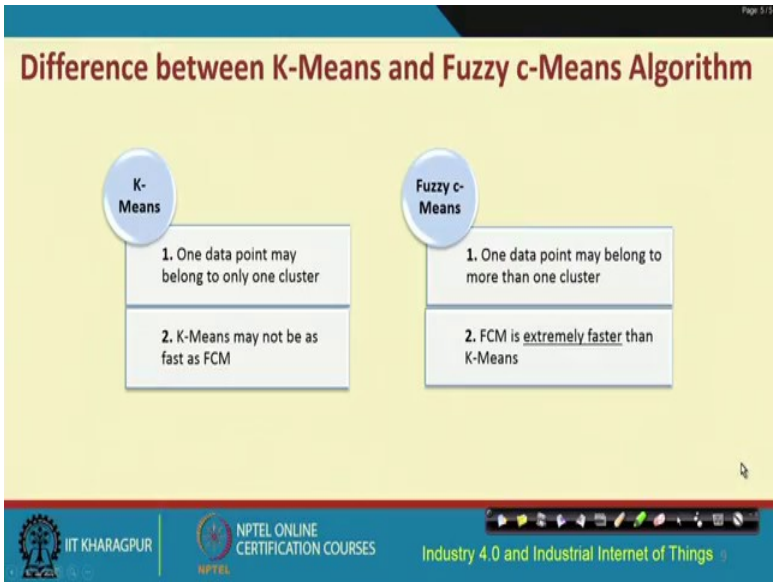
So, then what we do is we will in the next pass we will have like this and this one will be like this. So, what is essentially what has happened is that two clusters are formed based on their proximity to the centroid. So, two different clusters will be formed. Then, the next step you need to perform this one once again whatever we have done here needs to be performed once again.

So, we had these points, we again choose a centroid like this; we do exactly the same thing that we have done and we get a better centroid through this. So, we will have the renewed cluster means and thereafter we reassign the data points and we get these; the cross ones. So, we had started with five points. So, we have to have one more and same goes here as well.

We repeat the same thing we update the cluster centers and we repeat this process likewise and we reassign the data points and repeat this step in circles. So, how long do we repeat? We repeat until for two passes we get the same centroid. So, that is when this algorithm is going to converge. So, same or similar centroid if we are getting in two different passes of this algorithm when we are repeating this process. So, that is when this algorithm will stop execution.

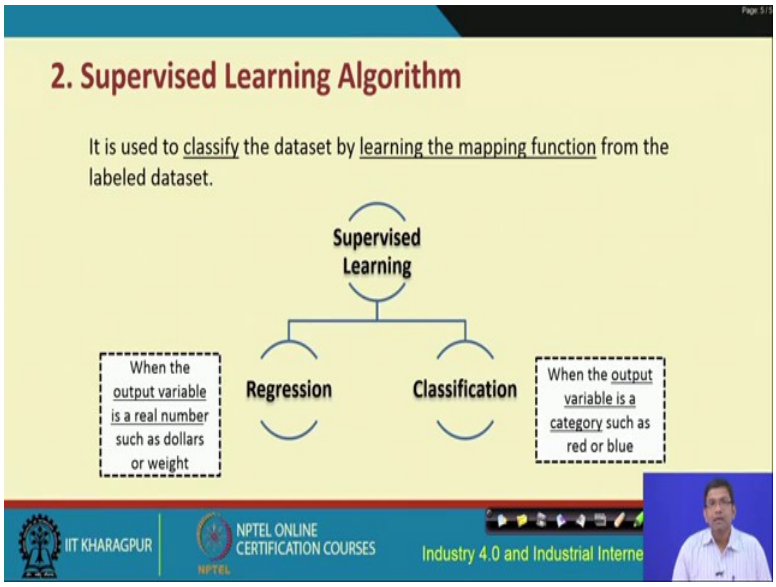
So, basically this is how this K-means unsupervised learning technique for clustering works. So, like this there are many different types of other algorithms that are also there in fuzzy actually what is happening is you do not have hard clustering like this. So, you do not have hard clustering. So, in Fuzzy c-means or FCM, so, what you are doing is that you may have because it is based on fuzzy logic one point can belong to both the clusters. So, that is how you will have some kind of a soft partitioning, soft clustering kind of approach and the same point can belong to multiple clusters in FCM. Unlike over here where you have definite distinct clusters to which one point is going to belong to.

(Refer Slide Time: 18:15)



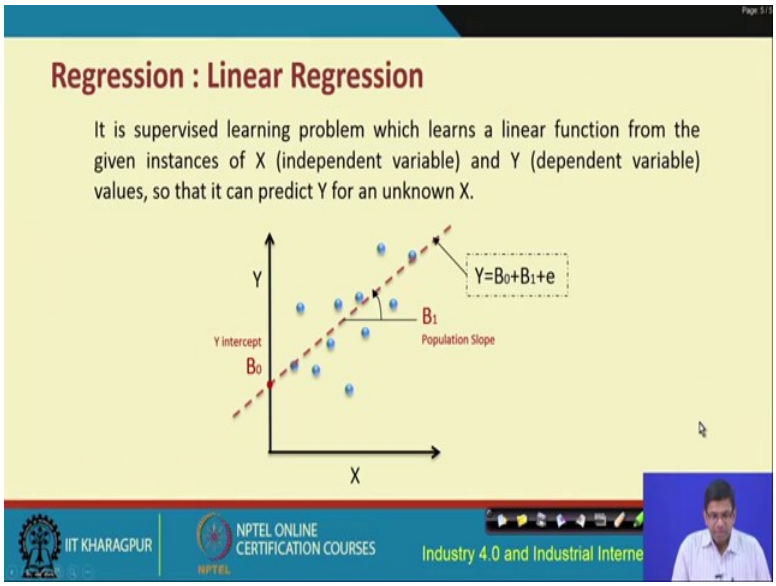
So, we will proceed further. So, this is this K-means where one data point may belong to only one cluster whereas in Fuzzy c-means one data point may belong to more than one cluster K-means algorithm is based on pure machine learning whereas, Fuzzy c-means is based on soft computing approach fuzzy logic. So, that is why it is known as Fuzzy c-means. So, because it is based on fuzzy logic FCM is extremely faster, much faster compared to the K-means algorithm.

(Refer Slide Time: 18:59)



So, next comes the supervised learning algorithm. So, in supervised learning algorithm, primarily we are talking about classification of data sets. Classification of data sets by learning the mapping function from the label data set. So, supervised, supervised means that there has to be some kind of label data set and based on the data set you are trying to make certain classification. So, supervised learning you know one technique is to basically classify the other one is to do regression. So, classification basically is when the output variable is a category; is a category such as red category or blue category whereas, regression is when the output variable is a real number such as the dollar values or the weight and so on.

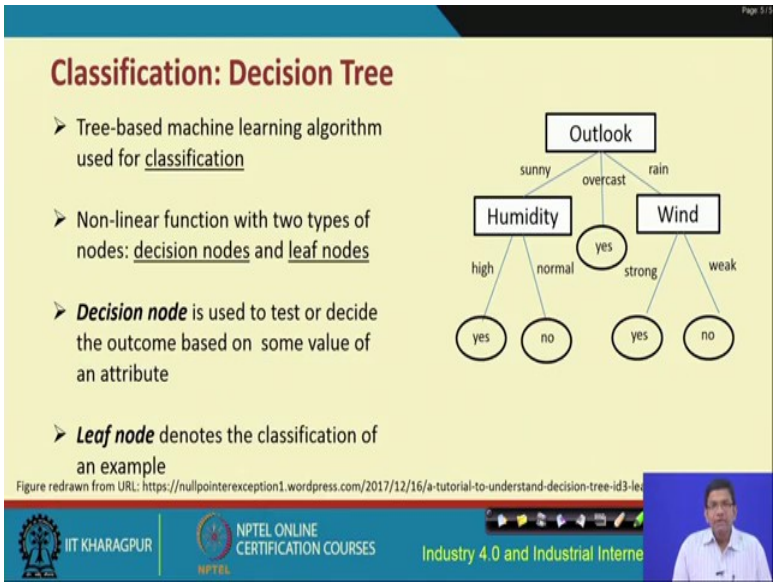
(Refer Slide Time: 19:51)



So, this is how this regression or linear regression looks like it is a supervised learning algorithm which learns a linear function from the given instances of the X and Y values, so that it can predict the Y for an unknown X. So, it is some kind of you know given a set of data points you want to basically fit line so that that will become the best fit .

So, that is the linear regression. So, that is the; we have all learnt in statistics also and that is the supervised learning technique and. So, basically this line over here. So, $Y = B_0 + B_1 + e$; B_0 is this intercept here, Y intercept; B_1 is basically this slope and e is basically the error. So, you have this kind of best fit kind of thing for linear regression.

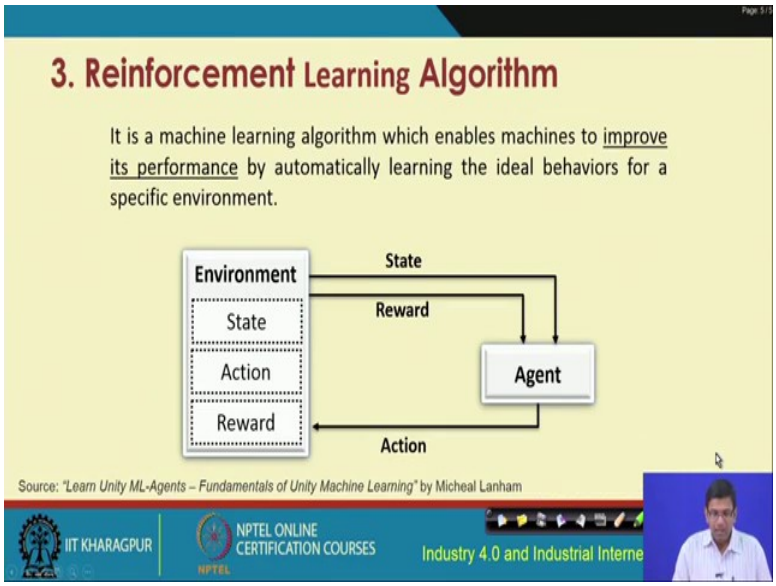
(Refer Slide Time: 20:53)



For classification techniques such as decision trees are quite common and decision trees basically as this figure suggests over here you are running some kind of a tree-based classification. And, so what is going to happen is there are two different types of nodes in this tree and these are basically the decision nodes and the leaf nodes. So, these nodes the white colored ones are the recent nodes and these nodes are the leaf nodes and so, this is a node basically will be used to test or decide the outcome based on some value of an attribute, whereas these leaf nodes will denote the classification of an example. So, this is how the decision tree looks like.

So, decision tree is also a very popular supervised learning technique.

(Refer Slide Time: 21:47)



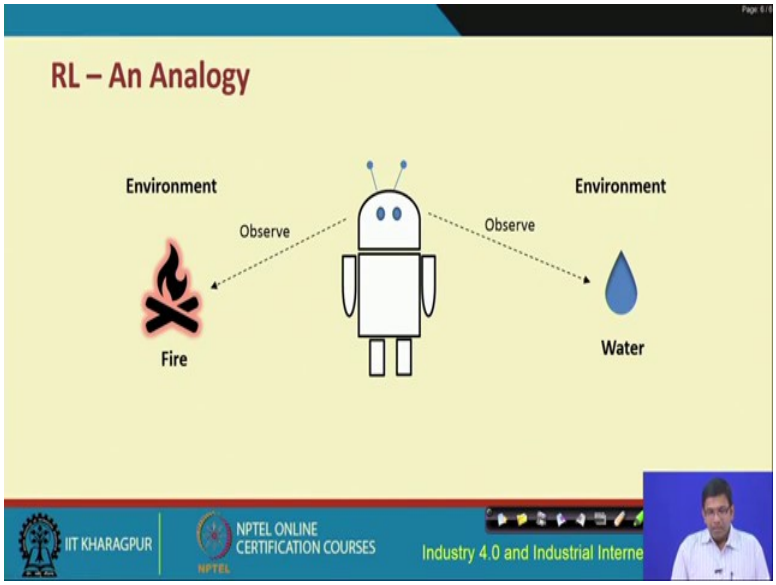
The third category is the reinforcement learning. So, it is basically a machine learning algorithm which enables machines to improve its performance by automatically learning the ideal behaviors for a specific environment. So, the way it proceeds is like this that we are going to have two different entities this agent which wants to learn and this environment on which or with which this agent interacts.

So, typically is going to happen is the agent has to learn by interacting with the environment. So, agent basically learns by interacting with the environment agent does not know that what is best action that it has to take. So, it starts like this that the agent will first take an action, and based on this particular action this environment basically is either going to reward or is going to penalize this particular agent for that chosen action.

So, a reward-penalty kind of mechanism, a feedback from the environment to the agent flows back and also the information about the change in the state is also fed back to the agent. So, with that information based on the reward-penalty value and the state information the agent makes its next choice and this is very important. So, you see even if we have a loop over here it is it we have to keep in mind in reinforcement learning that the next course of action that the agent will choose has to be dependent on these two; if that is not done then you do not have the reinforcement learning. So, sometimes that is a mistake that people commit.

So, so you have to choose the next course of action based on the reward-penalty and the state information.

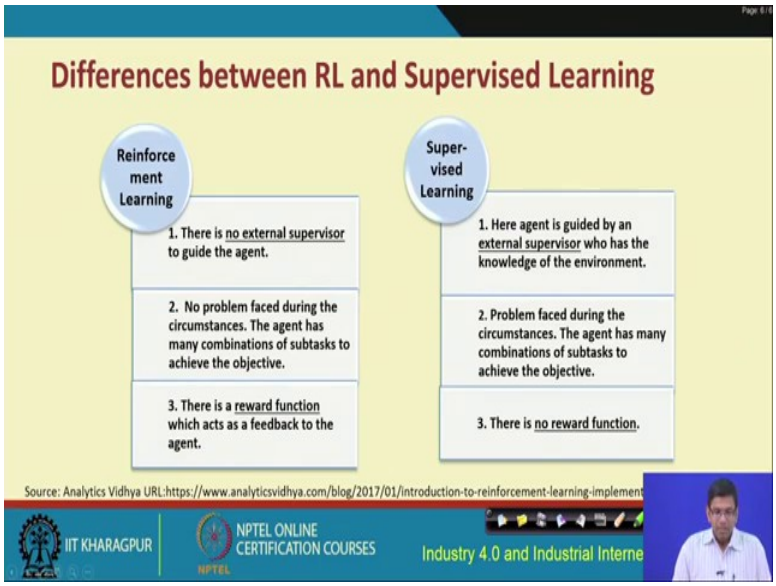
(Refer Slide Time: 23:59)



So, it is something like this that you have some kind of a robot or a human. So, in the same way as humans we learn through interactions. A robot also can do the same robot through it is interactions, observations, feedback from the environment and so on will know that what is what. So, basically the robot can through observations interactions with the environment robot will know that this is fire whereas; this is water.

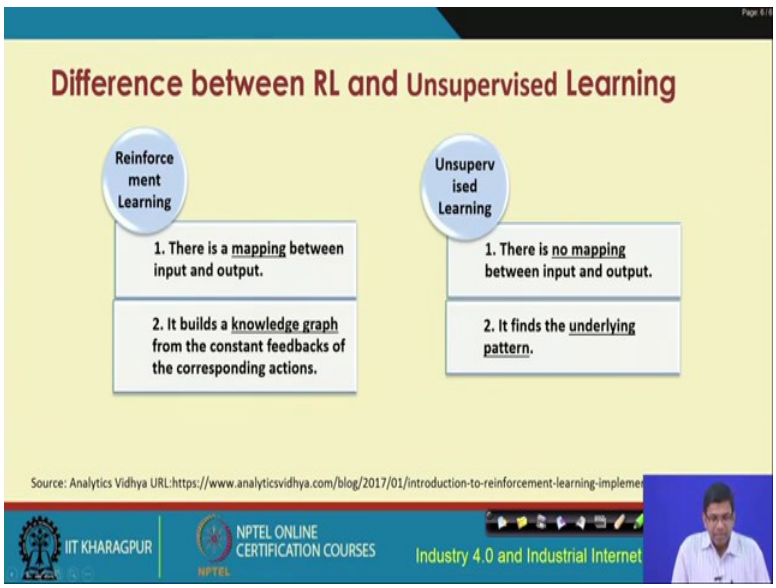
So, this is basically how this reinforcement learning in practice is going to work.

(Refer Slide Time: 24:47)



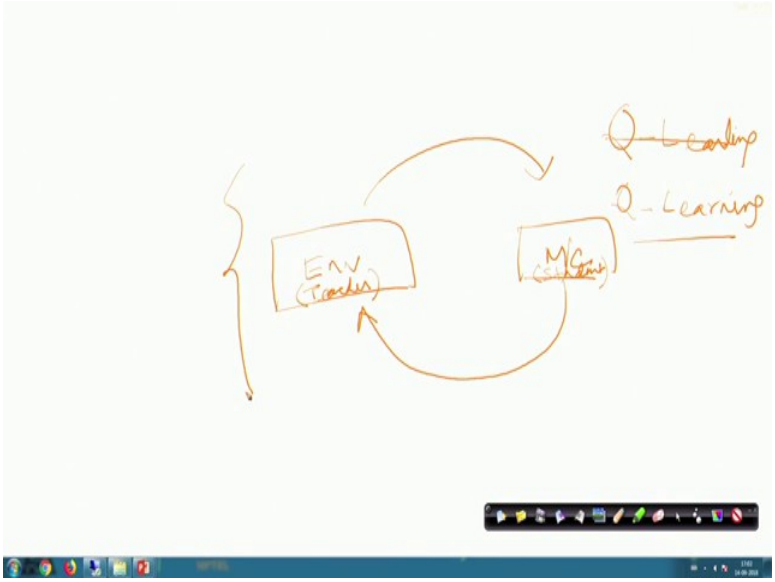
So, what are the differences between reinforcement learning, supervised learning and unsupervised learning? So, let us take up reinforcement and supervised learning first. So, in reinforcement learning there is no external supervisor whereas, in supervised learning there is some external supervisor who has the knowledge of the environment. So, that is where this training data set becomes useful. In reinforcement learning there is a reward-penalty structure that has to be in place whereas, because the external supervisor with previous knowledge is already used in supervised learning you do not need a reward function in supervised.

(Refer Slide Time: 25:35)



Comparison between the reinforcement learning and unsupervised; in reinforcement learning there is a mapping between the input and the output whereas, in unsupervised learning there is no such mapping. In reinforcement learning the agent basically builds a knowledge graph from the constant feed backs of the corresponding actions whereas, the unsupervised learning in that the agent finds the underlying pattern because here it is different. So, it basically tries to uncover the underlying pattern. So, that is the difference between the reinforcement learning and unsupervised learning.

(Refer Slide Time: 26:45)



So, basically what is happening in reinforcement learning is that I will give you an analogy in reinforcement learning what is happening is that you have a machine. You have a machine and you have this environment. This environment we can think of to be like a teacher whereas, this machine we can think of to be like a student.

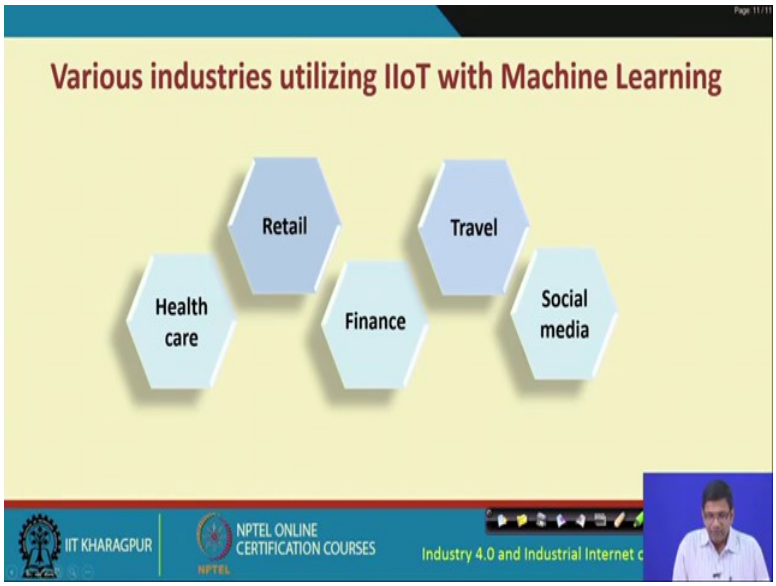
So, basically it is a learning kind of mechanism between the student and the teacher initially the student does not know anything, teacher knows. So, basically that knowledge has to be transferred to the student, but the student will not get that knowledge just like that. The student will have to ask questions; the teacher will say yes or no with a certain marks or penalty value and based on that the student will keep on interacting making different choices, student takes an action, the teacher asks let us say the teacher shows a color and the teacher asks that is it white then the student basically who does not know

whether it is white black or what whatever it is will say that it is grey with certain probability.

So, it chooses an action and the teacher basically will then teacher knows that no, it is not grey, but it is white then the teacher basically will penalize the student with a certain penalty probability. And, then based on that the student is again going to make a choice and then the student who is going to gradually converge towards the correct value by interacting with the teacher this is also known as Q-learning. There is a specific reinforcement learning mechanism which is known as Q-learning. This is more or less the overall idea and analogy with Q-learning. This is a specific type of reinforcement learning and this is how it works.

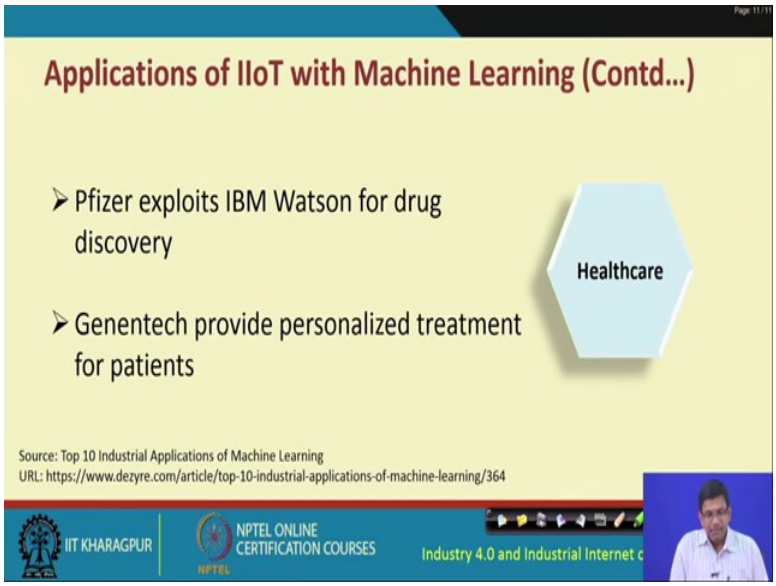
So, for IIoT there are different machine learning techniques in place.

(Refer Slide Time: 29:35)



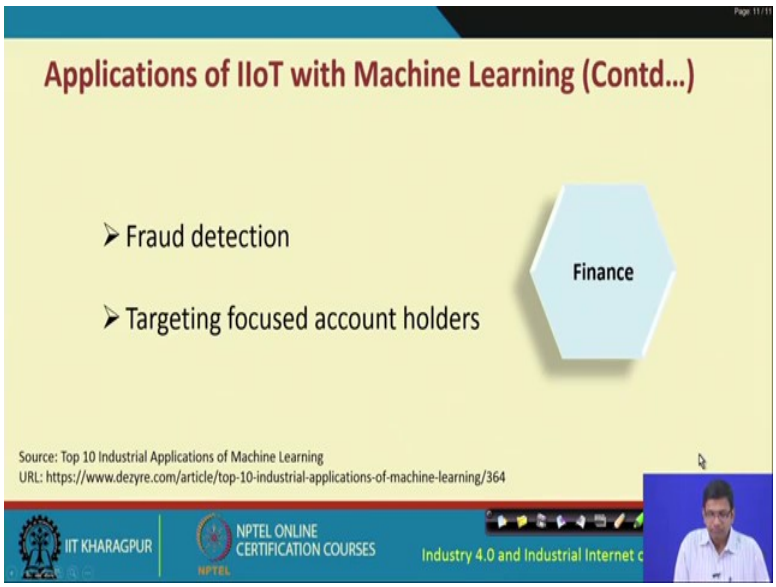
So, you could use machine learning in IIoT in order to harness the benefits, utilize the benefits of machine learning and make IIoT much more efficient and useful. Different industries such as healthcare industry, retail, finance, travel, social media and many more use IIoT with machine learning in order to improve their products their processes and so on.

(Refer Slide Time: 29:59)



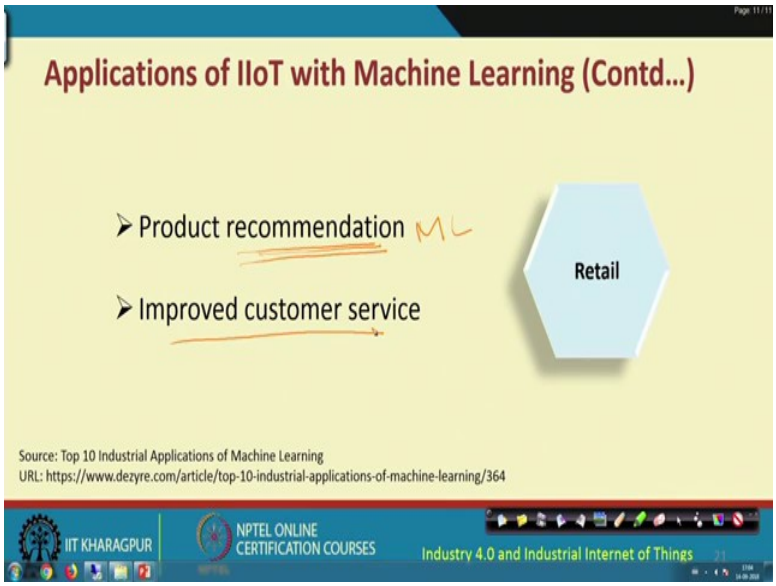
So, the company Pfizer which is in the medical space the healthcare domain they exploit IBM Watson for drug discovery. So, that is where they use different machine learning techniques for drug discovery. Another company Genentech provide personalized treatment for patients there also machine learning is used.

(Refer Slide Time: 30:23)



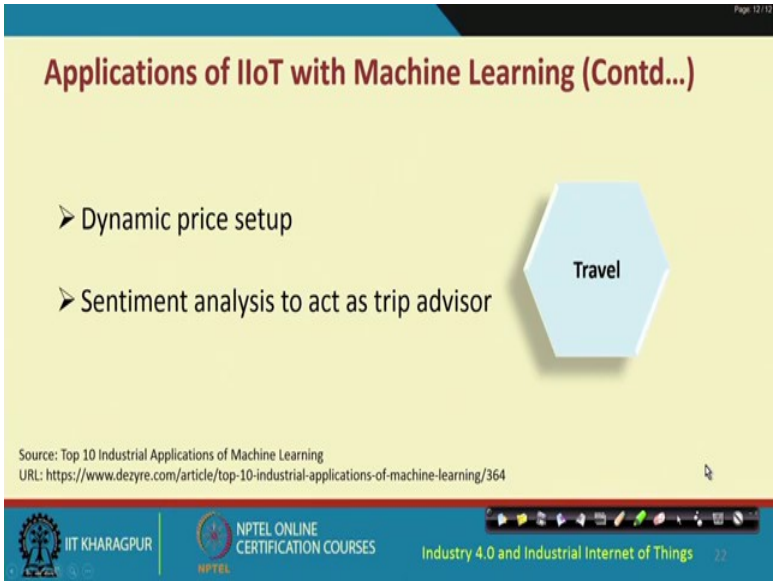
In finance, for fraud detection and for targeting focused account holders IIoT with machine learning techniques could be utilized.

(Refer Slide Time: 30:37)



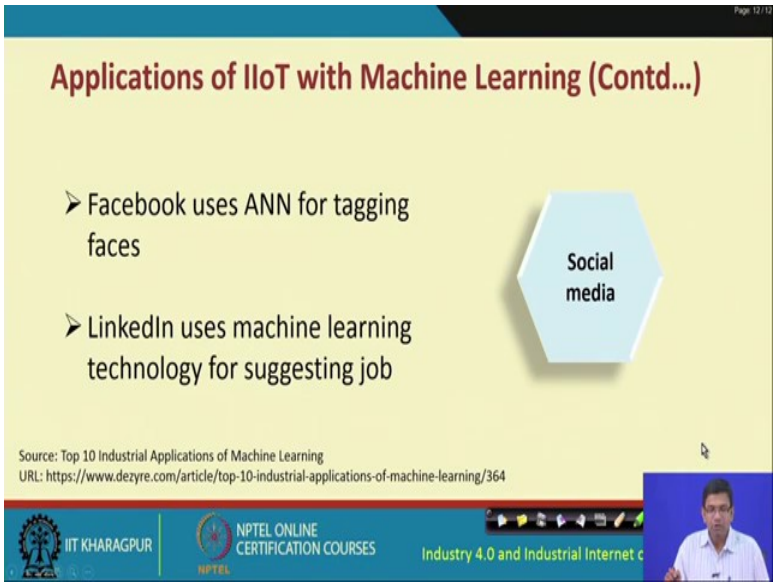
For retail, product recommendation, any recommendation basically this recommendation is something where recommender systems etc. are heavily based on machine learning. So, for product recommendation in the retail sector or for improving the customer services there also machine learning techniques could be used.

(Refer Slide Time: 31:09)



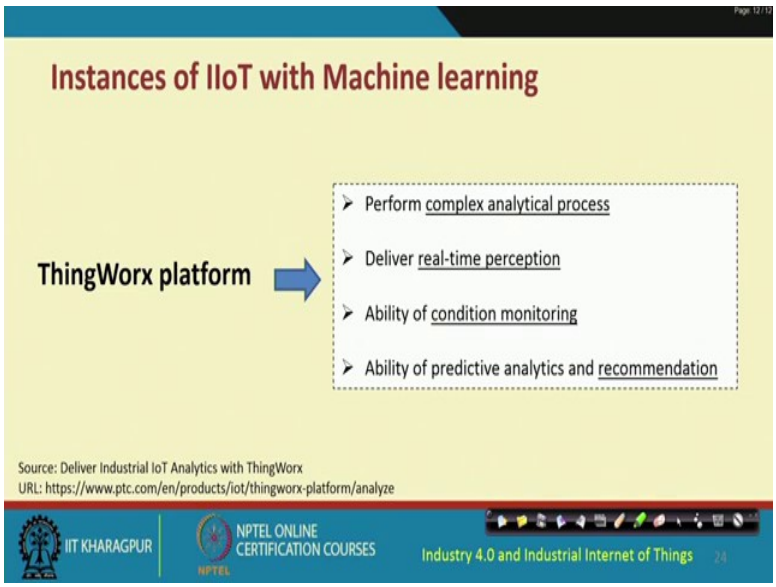
In the travel sector also IIoT with machine learning for dynamic price setup, for sentiment analysis to act as trip advisor IIoT with machine learning combined can be used.

(Refer Slide Time: 31:21)



For social media, Facebook uses artificial neural network for tagging different faces and this is what most of us have already experienced. So, behind the scene we see that Facebook basically does lot of these tagging, but that is based on ANN – artificial neural network which is a popular machine learning technique. LinkedIn uses machine learning technology for suggesting jobs.

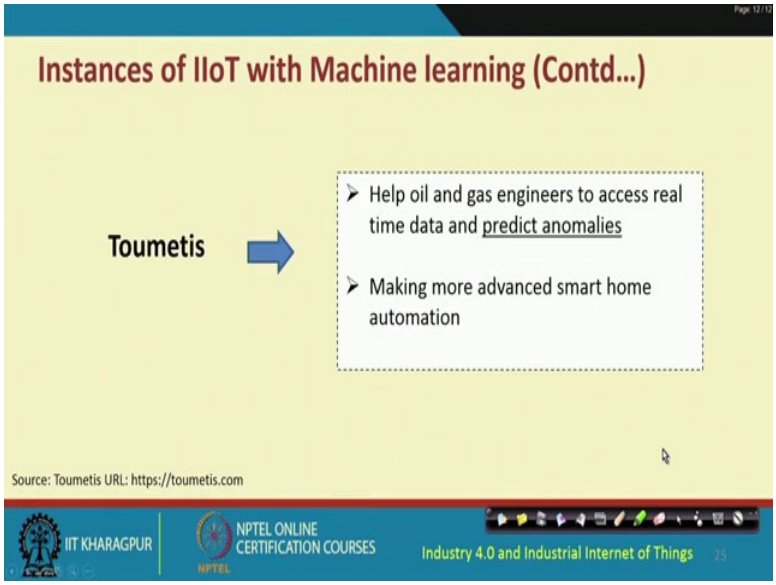
(Refer Slide Time: 31:51)



ThingWorx platform is another example which performs complex analytical processes, deliver real-time perception, offers the ability of condition monitoring, offers the ability

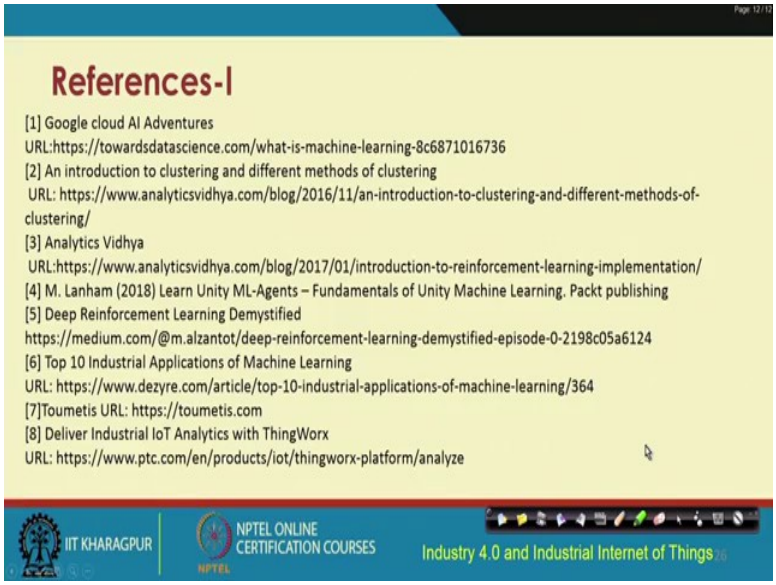
of predictive analytics and recommendation with the help of different machine learning techniques.

(Refer Slide Time: 32:11)



Another company Toumetis, they are in the oil and gas space. So, they help the oil and gas engineers to access real time data and predict anomalies.

(Refer Slide Time: 32:25)



So, these are some of these examples of the use of machine learning and often machine learning combined with IIoT. So, this is where machine learning combined with IIoT can be useful in different industrial settings and machine learning so far what we have

understood in this particular lecture is that machine learning can be of broadly three types. One is the supervised, unsupervised and reinforcement learning and there are different-different other machine learning techniques that are also there.

So, these are some of these different references that can help you to get a little bit more in-depth understanding of machine learning. If you are indeed interested to get more understanding of machine learning you are encouraged to go through some book on machine learning, but unless you have real curiosity and curiosity to know the different algorithms and methodologies that could be used in machine learning only this much of information should be sufficient for you. And, in case you are more interested you are encouraged to go through different literature and particularly the machine learning books.

With this we come to an end of the first part of lecture on machine learning introduction for IIoT.

Thank you.