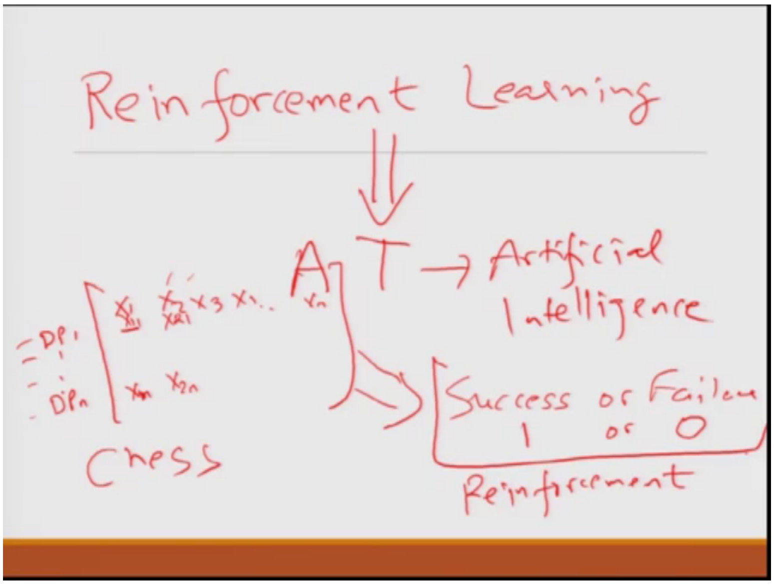
Practitioners Course in Descriptive, Predictive and Prescriptive Analytics
Prof. Deepu Philip
Dr. Amandeep Singh Oberoi
Mr. Sanjeev Newar
Department of Industrial and Management Engineering
Indian Institute of Technology, Kanpur
National Institute of Technology, Jalandhar

Lecture – 26
Machine Learning – (Part 3)

Welcome you all to another session on machine learning, in our course on descriptive prescriptive predictive analytics for practitioners. We talked about supervised and unsupervised learning in the previous sessions; we got an intuitive grasp of the same.

(Refer Slide Time: 00:40)



And today I will let me first begin with discussing another very popular form of machine learning, which is called reinforcement learning. In modern times currently if you go through the literature and go through the buzz around if reinforcement learning, this is also people the call it AI or artificial intelligence, though this was not exactly what was artificial intelligence always artificial intelligence is a much older field. But these days in many places when the people talk about artificial intelligence they are actually talking about reinforcement learning.

So, reinforcement learning is somewhere say between supervised and unsupervised learning, though its a different breed all together. So, what happens in reinforcement learning is that, you have one final output. So, for example, in supervised learning you have different data points, you have data point one, which are say if you talk of regression we have x 1, x 2, x 3, x 4 x n and then we have a dependent variable y.

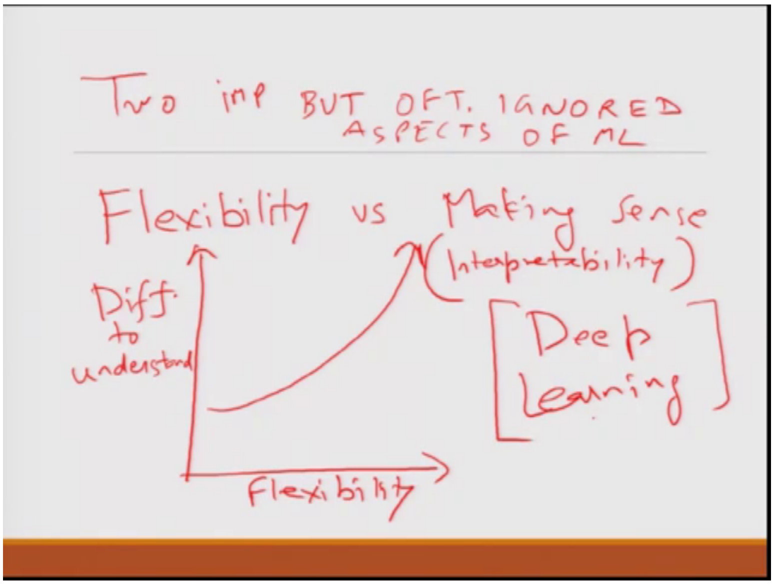
So, similarly you have DP n x 1, n x 2 n. So, this may be x 1 1, this is x 2 1. So, you have this multiple data points and for each data point you know what is the desired value. So, this may be y 1, this may be y n. So, the goal in supervised learning is to train the data such that the difference between this y 1 and the predicted y 1 that you get from the model is minimal. In reinforcement learning you do not have these instead what you have is a final output final output of success or failure 1 or 0.

So, just to give an example one of the examples may be just for say game of chess. So, in game of chess each of these different axes could be different moves that you make and after all the moves that you make as per the rules of the game, you get a final success or a final failure 1 or 0. So, the goal of reinforcement learning based algorithm is to optimize each of those moves each of those unit axes that led you to the final outcome. So, this 1 or 0 success or failure this is called reinforcement.

So, this comes from some classical theories of behavioral science that emerged in early 20th century late 19th century, if you would have heard about for example, Paulo's experiment and many such experiments. The whole idea was to give somebody a final feedback of success or failure and then hope that the organism, the human or even animals they adapt, they learn and then they optimize the system to create maximize the rate of success.

So, this is called reinforcement learning, in modern times if you want to study more about it you can read some new books, which take this topic all this topic artificial intelligence we will not get into details of this. Let me now move to another important topic, which I would say are very crucially even before you begin your machine learning or your model building exercise.

(Refer Slide Time: 05:06)



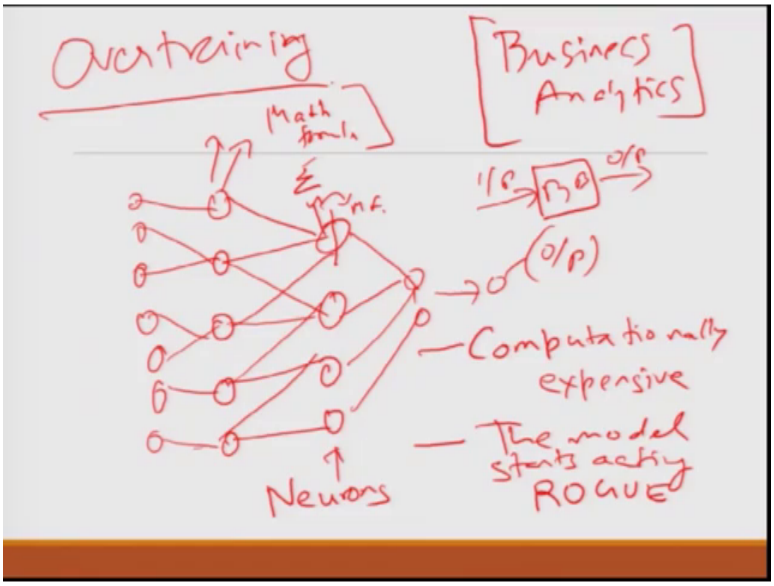
So, one is. So, these I would I would call them as 2 important, but often ignored aspects of machine learning.

So, the first is flexibility verses making sense. If I have to use a more technical word for this I may call it interpretability, which layman terms is trying to figure out what the hell is that am trying to do. So, if I have to roughly plot this on a graph, if I plot flexibility of the model verses difficult to understand. So, you will find that apart from very rare cases, its almost this curve which means that as you make the model more flexible. By flexible, I mean you have more parameters that adjust to give you the final results there are more ways in which the parameters interact and adjust with each other. So, you bring more and more flexibility in the model, but what happens is that the model becomes very difficult to understand.

So, flexible model is not always as you will see in the second ignored aspect of machine learning, but very often if it is very beautiful in the data on which you train the model and it gives you great result its looks very exciting and because there is a huge amount of now computational power, we have at our disposal it is often we are we are almost kind of inclined towards we are kind of I would say we have a natural attraction towards using more and more complex or so called flexible models.

Now, an extreme example is what is in fashion these days something called deep learning. So, deep learning is nothing, but fancy term for what used to be called neural networks some years ago, neural networks were very complex structures may be just to represent the over kind of these different nodes.

(Refer Slide Time: 08:10)



Each node would run some mathematical formulae and then it will pass the input to another set of nodes. So, there will be cross linkages you can define, and then in every node the first part will aggregate all this together and then it will be another mathematical function, then it will pass on and may be here there are multiple nodes and finally, one node and this will be the output.

So, you could play around with the mathematical formula, the mathematical functioning used. You could play around with the way you sum up the values you can normalize you can standardize you can do a lot of things; you can play around with the number of these nodes in neural network language called neurons. So, you can have multiple layers. So, this is just may be a 3 layers, you could have say another layer here. So, there could be multiple layers and deep learning is nothing, but trying to create more deeper more wider more networked more number of layers in will network and then trying to get as flexible as complex and output as possible.

The pit fall yes it definitely drastically increases your accuracy you can train the model to give really good results, but there are 2 major pit fall; number 1 it is computationally expensive and often in the era of cloud servers, you have to pay for a every bit of computation that you use. So, unless the results would justify something that we discussed earlier, they may sometimes not make sense the other part is that consider a scenario where you know the model starts acting rogue.

So, if you recall that movie of Rajnikanth that Robot. So, that Robot he made a very fantastic, robot and robot then started acting rogue it became a villain and often a case it has happened especially in business scenarios in specially in business analytics. Often what happens is that, organizations in their enthusiasm they go for something like a very complex model like this say deep learning for example, or earlier there was lot of enthusiasm for support vector machines, another set of complex algorithms and then the model could not run, after some time it would not give the desired result. The reason is that you have no way to control or understand what this linkages are, what is the logical reason behind this linkages that almost like a black box.

So, it is like a black box, input comes, output happens unlike this for example, if you look at a regression model in regression model you get the parameters which arte important, you get various coefficients of each of those parameters each of those terms in the equation, which give you the relative kind of a sense of how important this factor is like. You get your t statistics; you get a variety of things to understand which variables are important which are not and if a model is not working fine you can actually figure out and understand that may be in the new population the new data that I am getting certain variables are not important.

For example, you are running a marketing campaign and earlier age was an important variable now age is not coming important. So, may be the choice of the people have changed, the typically products liked by young the olds are slightly more conservative it takes time for old people to except a new innovative product or a new fun product cool product what they say the cool factor.

So, cool factor actually decays overtime and then. So, age may be an important variable for a cool product that you brought in today tomorrow it may not be there. So, all you have to do in a regression equation is to simply remove that variable may be re calibrate the model and you are done and often a times certain variables will not be important. So, even if that variable is not bringing result the rest of the variables will be able to hold the fort in case of a deep learning method because you do not even know what is happening sometimes it may over train.

So, there is a this concept of overtraining which will again come in a while in the next point what happens is that, you do not even know how it has trained itself and sometimes you get really weird results, as a classic example I think this was for sometimes done in US in military is said to have used a very state of the art neural network to identify certain tanks.

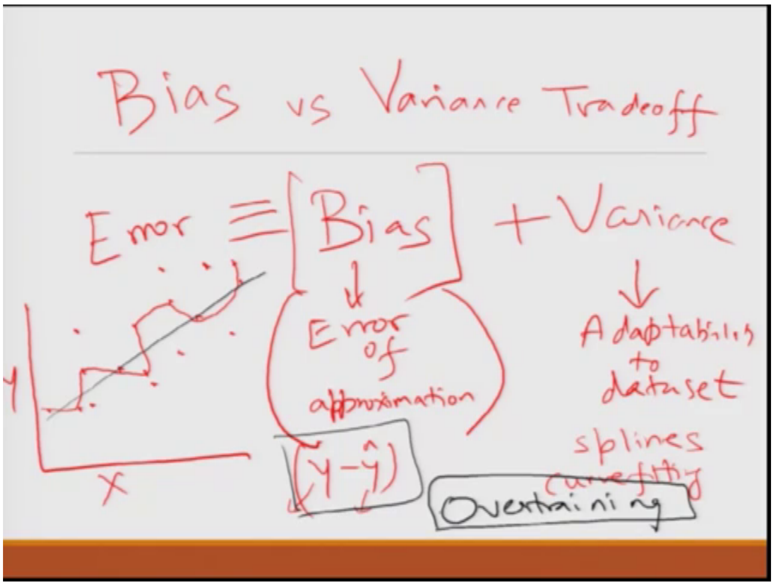
So, the photograph would there and they have to identify whether that particular object is a tank or not from a distance and it worked very beautifully, but then when they started to use it with the models started feeling very very I would say hopelessly. The reason was that what the model actually did was coincidently or all the photos all the images that they gave to the model to train it, were all tanks in day light. So, the model instead of training itself identify the tank, it remain itself unidentified day light. So, whenever a model of a tank came which was in night it would say it is not a tank.

So, because it was a neural network there are no factors nothing there is no way you could have figured it out unless you have done a very thorough testing and often in business situations what happens is that because the number of variables are so, huge because the number of spurious variables that can also crop in our so, huge that the chance of you missing something important or the model training itself on some of those really spurious things becomes really high.

So, always keep in mind and I think this is what we started in the beginning and let me again reemphasize that, unless you have a real valid reason to go for a complex model or a more flexible model. Simpler model a model that you can understand model that you can also think about, you can you can use your own intuition and business sense also around it might be more useful yes for things like image recognition, for things like videos, things like more non human data if I have to generalize and say.

For those kind of things definitely some of these advance models they would work fantastic for speech recognition, it will work fantastic this deep learning and advanced the more flexible model, but when you come to things like doing market research customer segmentation, thinks like cross cell up cell even understanding the behavior or the performance of your employees. So, these kind of human data management related data, they are unless you have a real valid reason. A simpler model with more thorough data analysis done data stewardship data cleaning done and all regress steps of model building, model validation adopted that would definitely apart from very rare cases would come useful.

(Refer Slide Time: 16:44)



The second important point I want to again mention and which is almost an extension of the previous in terms of its interpretation is bias verses variance tradeoff.

So, basically I will not get into the Math's of it, but the error the gap between the actual and the predicted values that you get from model, you are talking about supervised learning models out here. You can represent the error to be equivalent to bias plus variance. Now bias is in more intuitive sense error of approximation and variance is adaptability to data set.

So, what happens is more the flexibility of the model, more the complex model what happens is the error of approximation which is in simpler terms y minus y cap. So, when you try to minimize y being the actual output and y cap being the predicted output, when you try to minimize this totally and you remove the bias completely, the variance or in a way the unsystematic error I would say that actually increases.

So, you have taken care of the error you fit. So, as a, but may give an extreme example. So, you have this y verses x and there are for example, different data points. So, typically you can always create do a curve fitting and create a model which is almost like this. So, things like splines simple let us call it curve fitting. So, what you do is, you create model you create an equation which passes through all this points.

So, what happens is in the whole process you may feel that well once this is there now in this particular case, my difference of actual and predicted value is actually 0, but the moment you bring a new dataset. So, there is our new data points which comes which are further here, this model will fail to adapt to this new dataset while had you used say regression line, it would still work for this data sets in this case there would be this error, there would still be this bias, but the model will still work.

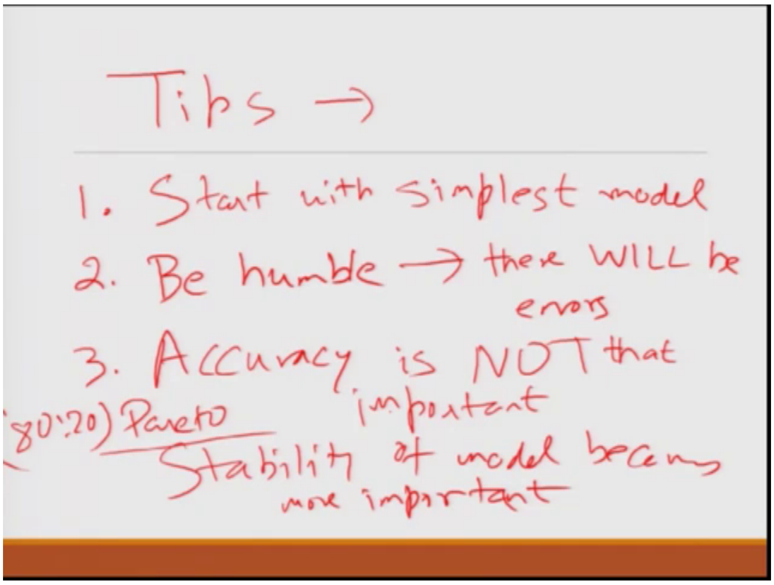
So, what is happening is that, more the more you adapt the model to a particular data set it loses its ability to predict the things with equal robustness in future. So, with a new data set, it is it is this is again this is something which is called overtraining. So, your model was so much trained in the existing set of data, that when a new set of data came it never dint have any clue what to do. So, such models.

So, typically when you create when you take use more complex models, these models will fit because there are more variables more parameters in the model, they fill fit very accurately with the existing data set. But then when you run the model on a new dataset the predictions often come out to be very poor and that is why you always have to strike a balance between bias and variance if you use a very simplistic model what happens is there is always this error of approximation.

So, for example, if you use a regression model, a regardless of whatever you do there is going to be some amount of error, because regression by definition says that it will try to create that assumes that the world is a straight line, it assumes that everything that deviate from a straight line in is an error point and it will try to minimize those errors.

So, even in the best of the case the best of the regression model will be still be some error because the whole model is defined by minimizing error and not making it 0, the least square distance least square error the way we call it in case of more complex models a very high degree polynomial model it may fit very accurately, but then the moment a new dataset comes the model fails you need to strike a balance from a practitioners stand point.

(Refer Slide Time: 22:34)



If I have to give you certain useful tips when you go on your machine learning journey, the number 1 would be start with simplest model. In many cases you do not even need to use machine learning, but simple descriptive analysis of the data some graphs some pie charts may also do the trick for you some tables.

So, start with the simplest often try regression and then incrementally improve the complexity of the model and for every time when you increase the complexity of the model, try to understand whether the cost of this complexity justified by better accuracy or better stability. And when you build the models again be humble, there will be errors. Often in new enthusiastic machine learning programmers I see this urge to try to eliminate error and one of the simplest things as we discussed in the previous slide, that you know if you try to over fit and try to remove the error totally that may actually be a blunder.

So, understand there will be an error because you never get accurate data there is always noise in the system, your goal is not to eliminate error, but to manage error. And here comes especially if you are from a management field, if you are doing this analysis for say field of finance, field of marketing, field of human resource please know that accuracy is not that important, just said it is important most of the literature in machine learning most of the literature in data analytics in general is all about trying to improve the accuracy, but there is always that hidden message and that is why we have so, many complex techniques.

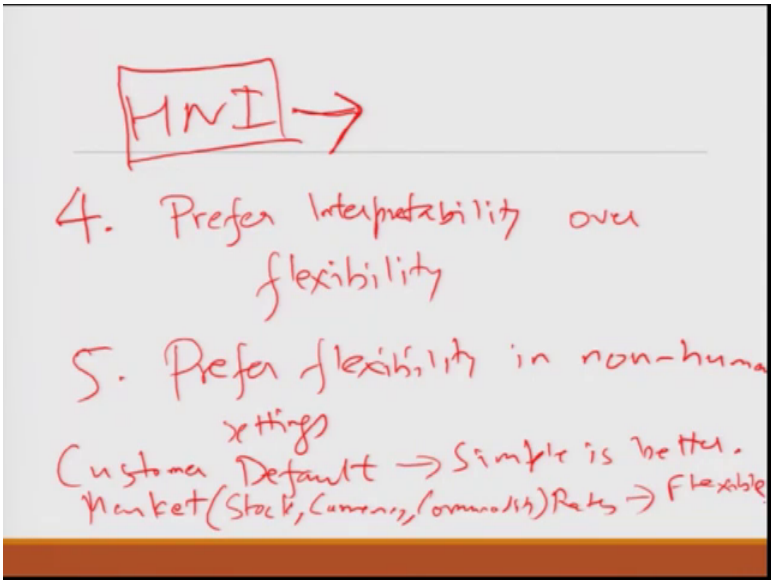
So, we have to increase accuracy and remove errors as I said curve fitting may be the best way given the computational powers we have. Accuracy is the hidden message is always that yes accuracy is important, but accuracy is not everything. Specially if the data is noisy the data is coming from sources, which are not designed to provide data with infinite precision more so, say human data, in that case more than accuracy, stability of model becomes more important.

And this has lot of repulsion because in an in a in a enterprise setting a very accurate model that dies after some months you have to again rebuild the model. So, it is always better to have a model which is more stable which can run with you for a longer period than work with a model which is very accurate and will die. So, you have a choice, whether you wants to keep building model there is a price to building a model there is a price to maintaining the model there is a whole number of steps that you need to do before you bring a model to production.

So, whether you want an accuracy or give a will in to compromise that accuracy for stability. So, you have to find. So, often you will see that that 80 20 Pareto rule that is more. So, valid in case of machine learning models you will find that well the incremental benefit, that you get in many cases beyond say a regression or a logistic regression model a simple linear model is just may be 5 10 percentages of accuracy improvements, but do you want that accuracy for a model that will die closely is the price the value that will generate.

So, high may be maybe not. So, yes often what is done many a times for different kinds of segments of people, people build different models. So, if you have for example, you are making a model for your customer management.

(Refer Slide Time: 27:38)



So, for the MI segment, there you know that the price of you know spending every dollar that you spend on treating this customers as VIPs and nurturing them is well paid off, you may go for more accurate more robust models because its also a dataset that you can manage, you have individual who are also personalized like the relations. But if you are going for model for the normal population for example, in your if your in the business of credit card.

So, if you have a credit card whose annual rate is say 1 lakh rupees or something you know that is only the celebrity were going to take that card. So, for those cards a kind of models that you build, they can be more accurate models. For rest of the population you may well do with more general models, which will be stable run over type and again the (Refer Time: 28:37) they are also that if these are such high network individuals and whom we are giving your individual, attention then you actually do not need that amount of model.

You need more of time, more of ways to convert the human judgment of people in relation with them in to something more sensible then you using all he Math's to target this people because in any case these relationships have to be very very customized.

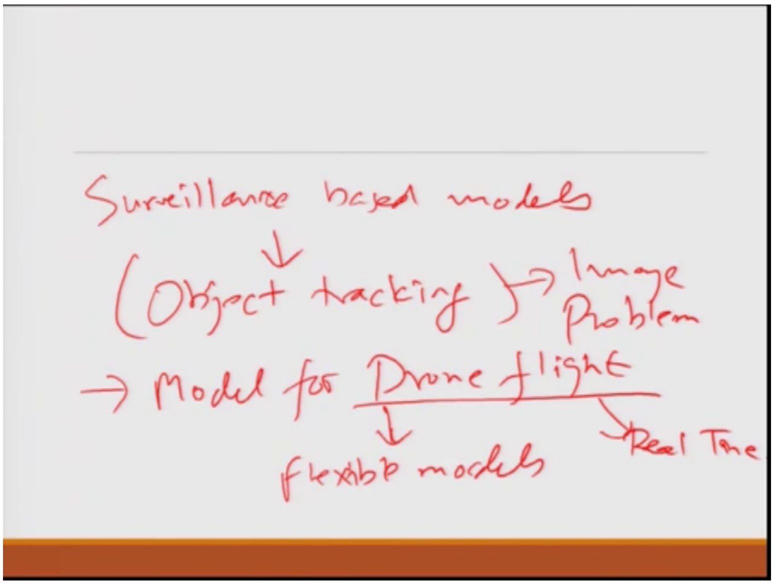
So, if you ask me in a business setting, there are not many situations where you would prefer to go for really complex models, you may do it for a hobby you can do it as a secondary models to may be test certain things out, but in most cases simpler would be best. And that brings me to the fourth point which is unless you have a reason; in most of the business settings you always prefer interpretability over flexibility.

You need to have models which you can intuitively understand which you can make sense about, many of the top notch analytics forms whose entire business zone analytical modeling and who have the server capacity to run the most complex models, they still work on simple models like regression in most of the cases. So, simple reason is that that helps there senior management to also understand what the model is doing, tweak the models when situation suddenly change the economy suddenly changes.

It also helps them play around experiment with new kinds of products and stuff and make sure that the model is slave and they are not being slaves to the model. However, if you are using a model for something like graphics images, different kinds of patterns, reliabilities estimates for machines and stuff they are probably more flexible models, the non human settings there these flexible models may be more relevant.

So, some let me. So, just for example, give some examples like if your modeling customer default, I would say simple is better. If you are trying to model market I mean stock currency, commodity rates, there probably you can play around with more flexible models though I doubt and this is again a an area of debate.

(Refer Slide Time: 32:18)



If you are using say surveillance based models object tracking. So, for example, model to track presence of any spurious person around your secured area. So, more often this is more often image problem; often high security zones they use it or you want to make model for drone flight there you can use more flexible models; however, in this case you will have to make sure that the model is smartly embedded and each time you do not have to because, these have to be more of real time decisions and hence you cannot afford to have models where the calculation of the output itself takes time.

So, the model building may be a very complex process you can play around, but then finally, comes you build the model, that model should be something which is very quick to execute. So, depending on this on different types of situations for voice recognition for image recognition two are the classic examples, where this deep learning is useful.

May be for behavior of the patterns understanding patterns of success failure in casinos that its very deep learning may be very useful, but something more down to were something more human, human resource performance evaluations stuff like that definitely go for simple model. So, we covered the broad essence of machine learning today, in the next session we will go through the entire gramat of machine learning algorithm, we will also talk about certain important things that you need to take care of to make sure your models are not only built well, but they are also validated. So, as that you can sleep with peace that the model will run with stability or any new dataset so.

Thank you very much.