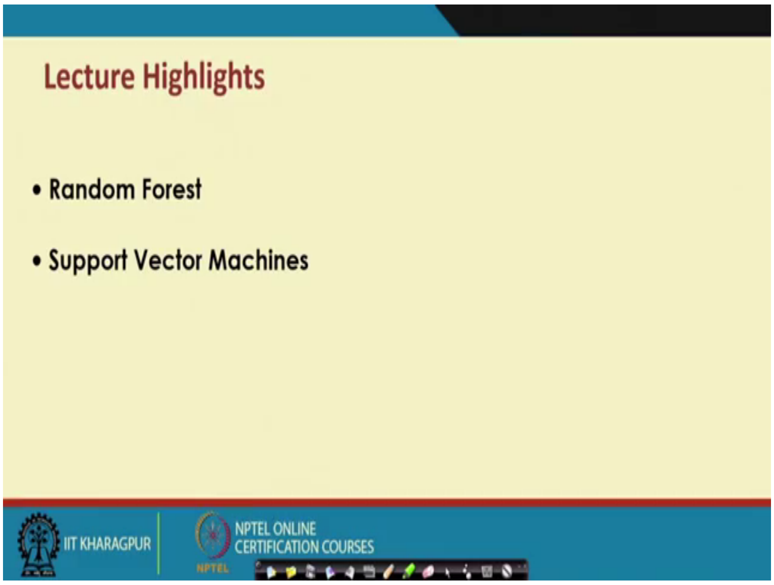
Business Analytics for Management Decision
Prof. Rudra P Pradhan
Vinod Gupta School of Management
Indian Institute of Technology, Kharagpur

Lecture – 38
Predictive Analytics: Machine Learning (Contd.)

Hello everybody this is Rudra Pradhan here. Welcome to b m d lecture series, and today we will continue with predictive analytics and that to coverage on machine learning. In fact, we have already discussed this particular component in the last lecture, and the topic which we have discussed earlier was artificial neural network, and in this lectures, we will be continue with the similar kind of you know technique, by covering two more additional kind of you know components, that is on random forests and support vector machine. All these are predicting predictive techniques predictive analytics techniques, and we need to again predict a particular variable dependent variable subject, to certain independent variables.

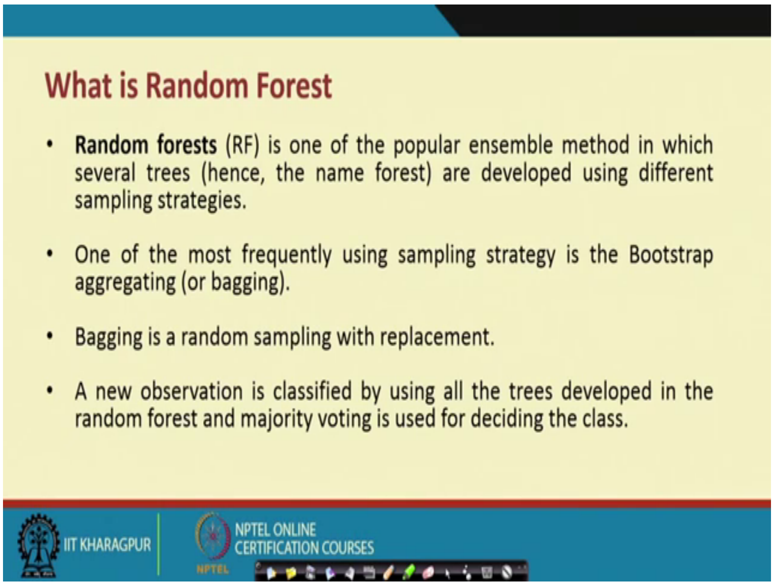
So, these techniques are basically classification techniques, and through which we need actually for business problems and, in fact, in reality lots of business problems are there. So, we need you know kind of you know classification, to understand the problems, and to predict the problems as per the particular, you know business requirement, or you know management requirement and a random forest, you know random forest and support vector machines are you know such techniques, which can help lot to do the classification to understand the data, get the insights, and then you know predict the kind of inner business problems as per the particular you know requirement.

(Refer Slide Time: 02:00)



So, in this particular you know lecture we first start with you know coverage of you know random forest, and then we will follow with you know support vector machines.

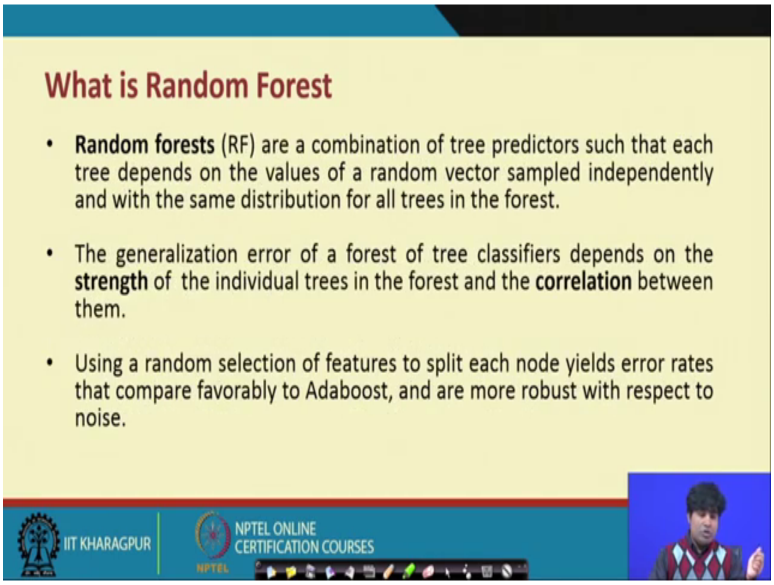
(Refer Slide Time: 02:06)



So, random forest basically you know a kind of you know method in which, several trees are developed using different sampling strategy. One of the most frequently using sampling strategy is the bootstrap aggregating, and that is called as you know bagging. And in fact, this particular technique is followed by a random sampling that to with replacement, a new observation is classified by using all the trees developed in the random forest, and majority voting is used for deciding the class.

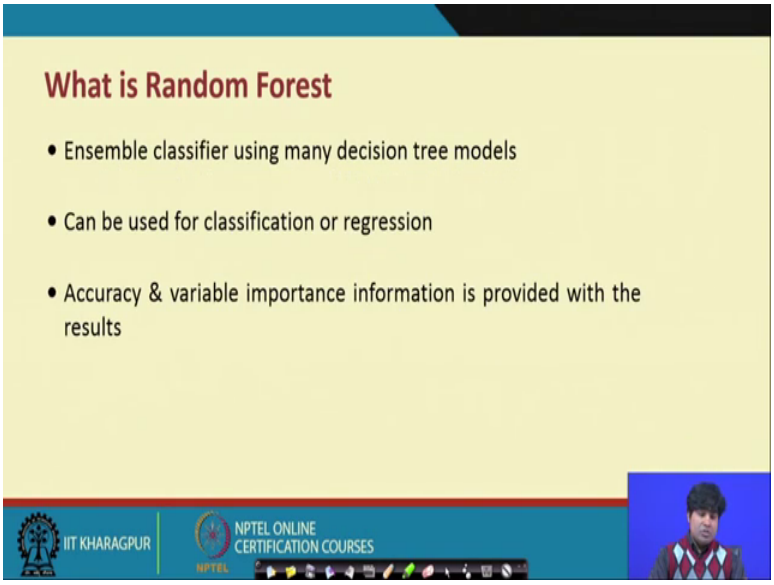
So; that means, actually we have a particular you know data set, with respect to you know several variables, and like artificial neural network will have also structure like, you know trendy particular, you know structures and then develop a kind of you know a classification or develop a kind of you know structure through which, we do the similar kind of you know predictions as per the particular you know business requirement.

(Refer Slide Time: 03:15)



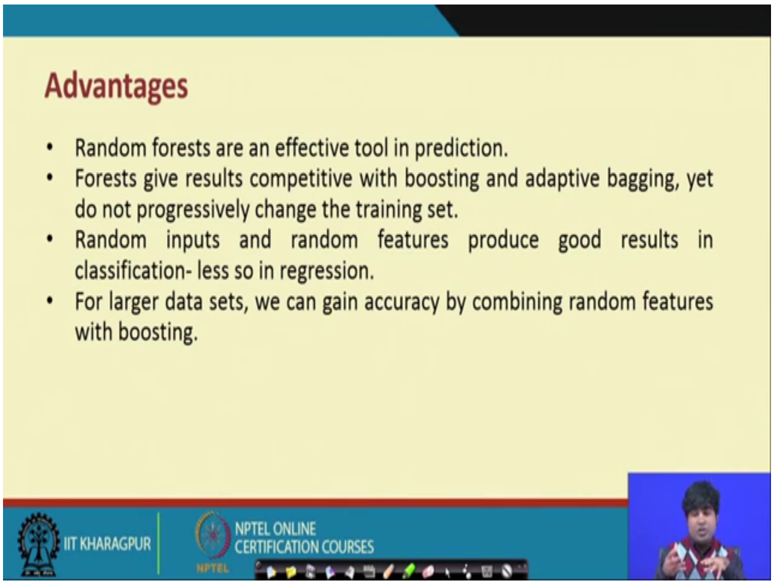
So, technically random forests are the combination of tree predictors such that, each tree depends on the values of the random vector sample independently, and with the same distribution for all trees in the forest. So, the generalization error of a particular you know forest of tree that is the classifiers, depends on the strength of the individual tree in the forest and the correlation between them. In fact, using a random selection of futures to split, each nodules, error rates that compare favourable to a adaboost and are more robust with respect to noise.

(Refer Slide Time: 03:59)



So, in typically means typically in total this is a kind of you know classifier, using many decision tree models. We have a kind of you know component called as you know decision tree, and the random forest is a kind of you know tool it is well connected with you know decision tree, and through decision tree will do the kind of you know classifications, and then we will go for the kind of you know prediction as per the management requirement. And it can be classified you know classified as per the particular you know requirement so; that means, you know the classifications like you know the concept ecology no cat, then the accuracy and variables importance information is provided with this particular you know results.

(Refer Slide Time: 04:51)

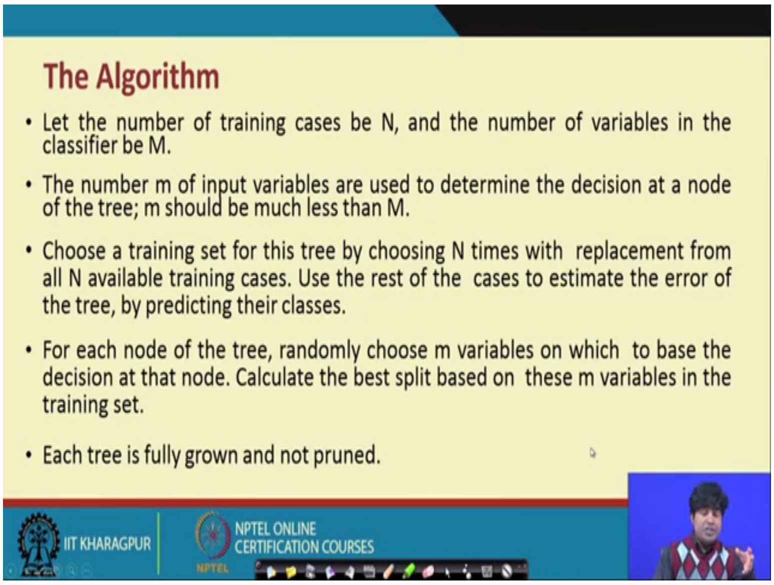


So, the kind of you know requirement is to understand the particular you know structures, and do the kind of you know forecasting as per the a problem requirement. So, we have lots of you know advantages, random forests are an effective tool in prediction, and forests give results competitive with boostings and adaptive bagging, and it is a kind of you know flexible kind of you know structures.

So, when will it change the particular you know training structure or the kind of you know data structures then; obviously, the particular classification can change, again as per the business requirement. So, random inputs and random features produce good results in classifications, and means that is technically compared like you know regression structures. For the larger data sets we can gain accuracy by combining, random features with you know boosting.

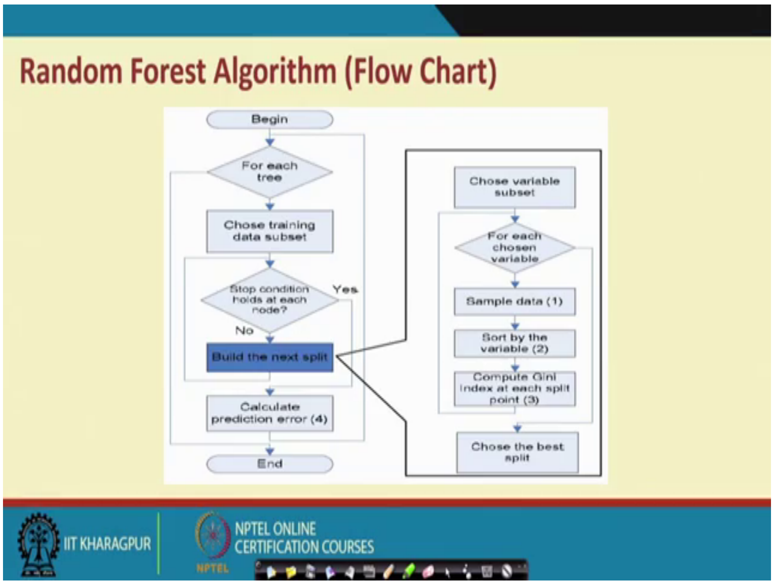
So, like you know artificial neural networks. So, in this case also we can do better predictions, and better forecastings, better classifications provided. We must have a several kind of you know data points through which, we have a lots of you know flexibility to train the particular you know structure, and come with a particular you know set up through which will you do the better predictions.

(Refer Slide Time: 06:08)



So, the algorithm is like this the random for you know random forest algorithm.

(Refer Slide Time: 06:12)



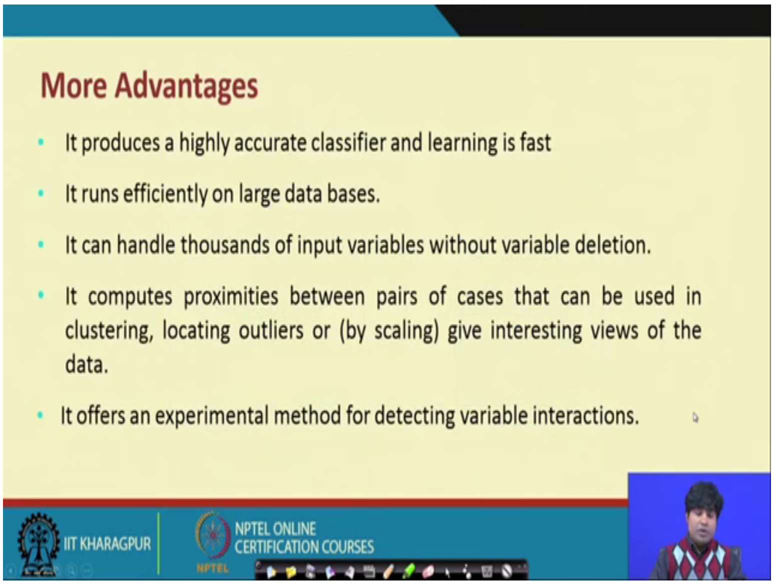
So, it we start with you know number of training cases N, and the number of variables in the classifiers you know can be denoted as a capital N, and then we pick up a kind of you know N input kind of you know variables, which can you used to determine the decision of a particular you know requirement, and that too for a particular you know tree, and where m should be less than 2 m. So; that means, technically we have a larger set and within the larger set, we will find out a small cluster through which you do the classification as per the particular you know requirement so; obviously, the particular you know training structure, the prediction structure, followed by a random samplings and that too with replacement so; that means, we have a plenty of you know flexibility to change the structures, the you know train the kind of you know data set, then to come a you know come with a particular you know setup through which we can do the predictions as per the, you know problem requirement.

So now the particular flowchart can give you the details about this random forest. So, we start with you know developing a tree that is a decision tree, and through that decision tree we will it do the kind of you know structuring, and you know create a kind of you know environment for prediction. And here to start with we have a kind of you know structures for each tree, choose training data subsets, and then we continuously you know follow up you know like a you know depending upon the particular you know prediction structure, and you know errors.

So, continuously we can actually train and you know structure restructures, you know when the when it is not coming as per the particular requirement, then against it will repeat the particular you know proceed process. And then finally, we come with a kind of you know structure through which really do the kind of you know predictions.

So, the mechanism says sometimes in this process is called as you know gini index, and that will help you again to you know prepare a kind of you know structure through, which you will do the better predictions. So, it is actually kind of you know flexible systems, or you know through which you will develop a particular you know structure or setup, through which we can do the predictions.

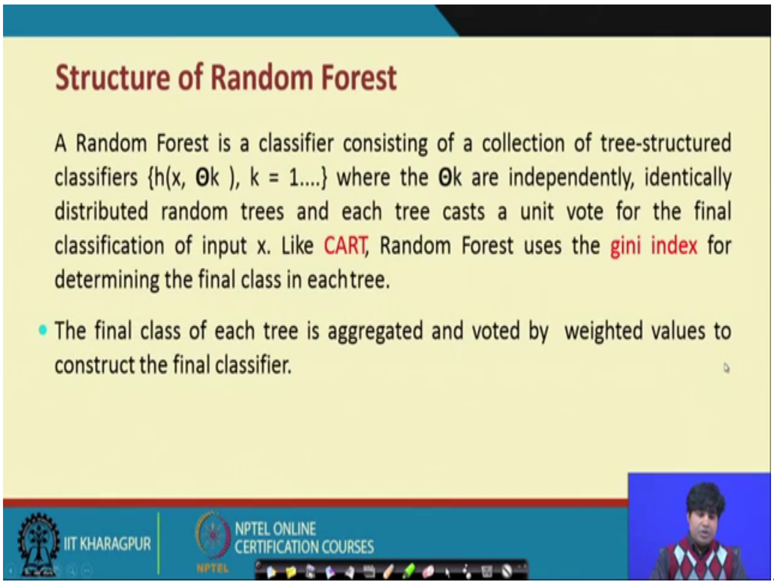
(Refer Slide Time: 08:38)



And we have lots of you know extra advantages to use this random forest, and it is highly accurate classifiers, and learning is very fast like you know in the case of in on neural network and. In fact, as usual the if you compare with the simple regressions, this particular mechanism is more attractive because it is having actually high accuracy to develop the kind of you know classifiers, and then turn the kind of you know structure through which will do the better predictions.

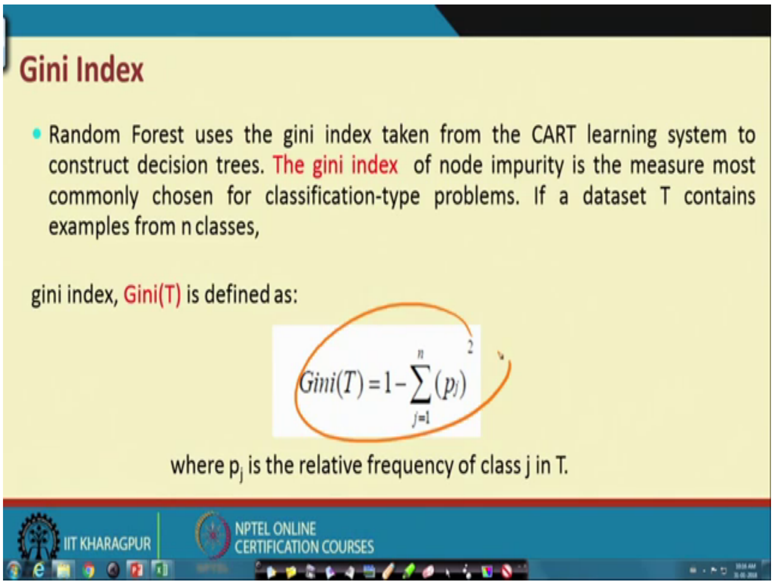
So, ultimately compared to simple regressions, in this kind of you know technique like random forest, support vector machines and the neural networks every times we must have been in a large data set, through which you can do better training develop a better structure, and then we will go for the kind of you know forcastings as per the kind of you know requirement.

(Refer Slide Time: 09:29)



The structure of the random forest is like this. So, the kind of you know structure is called as you know classifier. So, we have to develop a kind of you know, structure through which we can actually you know develop a kind, of you know system through which you can continuously change the trains structures, or the kind of you know set up through, which we will get the best result and of course, that will be a check through the kind of new production error, and the system which we like to follow is called as a classification and regression tree, through which you know random forests can even develop the system through which you can do the kind of you know management decisions..

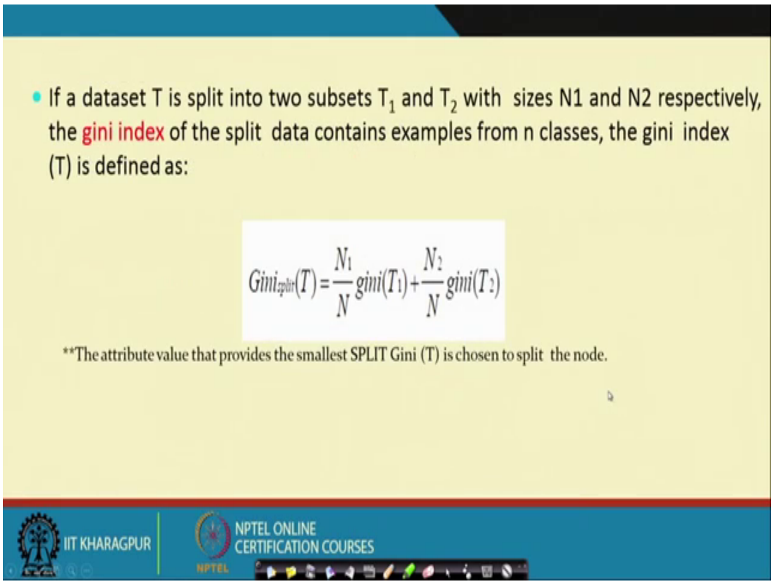
(Refer Slide Time: 10:17)



It in fact, it is a kind of you know structure like you know neural network, and through which you will develop the kind of you know set up, and we will go for the kind of you know predictions. So, what I have mentioned that you know the in the random forests the typical component, which you know we simplify the particular structure, is called you know gini index.

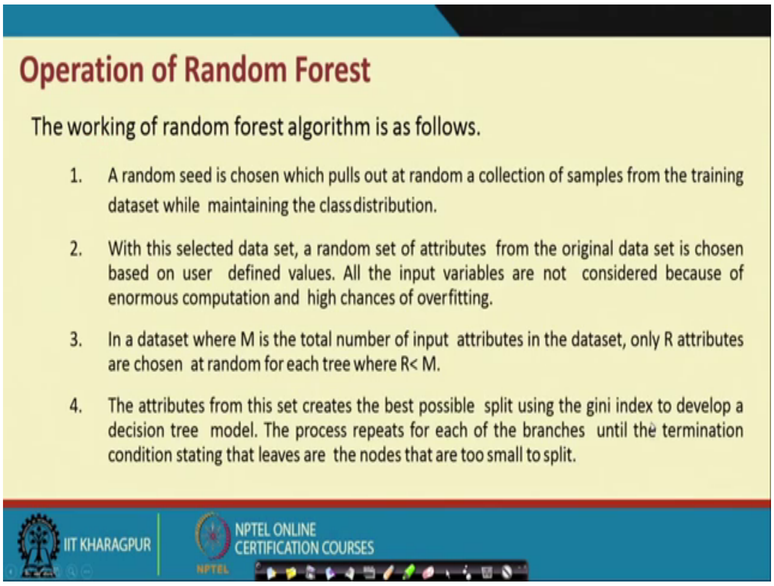
And it is actually a developed through his this particular you know structure and then I think you know this is this is the kind of you know formula through which you can calculate the gini index, and p is the kind of you know relative frequency of a particular class a z class, in a kind of you know T observations and then we accordingly actually predict the kind of you know structures.

(Refer Slide Time: 11:06)



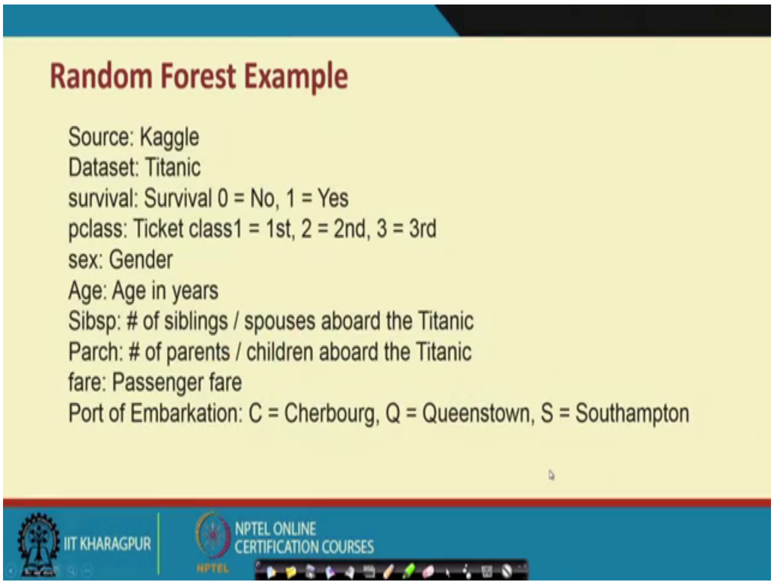
So, likewise actually we can we can divide the data set into two different parts like you know T 1 and T 2, and against we can create a separate gini index, for T 1 and T 2 and finally, the combined gini index is nothing but, you know gini index for the part ones, that is this subset 1 and the subset 2 and then finally, we can actually go for the kind of you know predictions.

(Refer Slide Time: 11:27)



So, you know. So, likewise we will you know develop a kind of you know structure, through which random forest can help you to predict the kind of you know business problems, as per the particular you know requirement so; obviously, you know we have a kind of you know systems, and the algorithm which I have mentioned here, that you know it is a kind of you know random sampling with you know replacement, and then we have a pool of you know variables through which the training structures, we pick up a particular you know structure or you know set, through which you can develop a kind of you know framework and with the particular framework we will go for the kind of you know predictions.

(Refer Slide Time: 12:09)



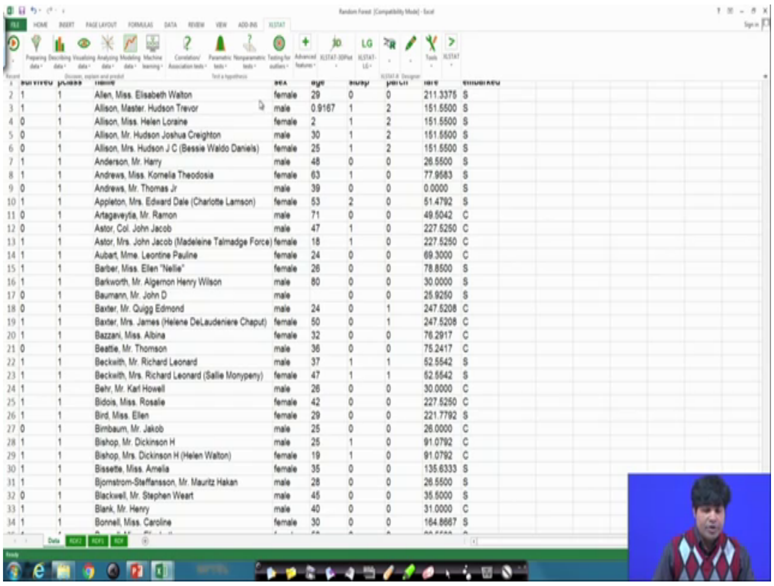
So; obviously, this is a kind of you know structure through which you can develop. And to understand the particular you know setup how random forest can help you to do the classification do the kind of you know prediction. So, we can start with simple examples, and this is actually titanic examples. And so, in the case of you know. So, we have a couple of you know variables in these particular problems. So, we have a variables both in qualitative types and the numeric type.

For instance, the age is a kind of you know quantitative variables, then sex is a kind of you know qualitative variables, then class that is actually categorical variables, similarly survivals it is also categorical variables, and likewise we have fares that is actually quantitative variables, then the port there are 3 ports which you have actually designated as a C Q and S.

So, likewise we have actually couple of variables through which we will do the kind of you know classification and the kind of you know predictions. So, in this particular you know problems. So, our dependent variable will be the kind of you know p class, and then it will be predict predicted that is, p class means that is the ticket class first second and third. So, and it is a kind of you know categorical. So, denoting 1,2,3 and then. So, that will be classified and we will do the kind of you know predictions, with respect to couple of you know dependent variable, and independent variables which you have already highlighted here.

So now with the particular you know information. So, I will take you to the kind of you know software's, and through which we can actually we can get to solve this particular you know problems. Usually this kind of you know structures manually you cannot do.

(Refer Slide Time: 13:56)



So, the particular software which you can actually apply to solve this particular you know problems through random forests is excel start. In fact, we can also do through r software's, but this is a since we are you know solving so many problems through excel data structure. So, again so, excel stata excel start can help you lot to solve this particular you know, problem by the use of you know random forest. So, in the excel start software.

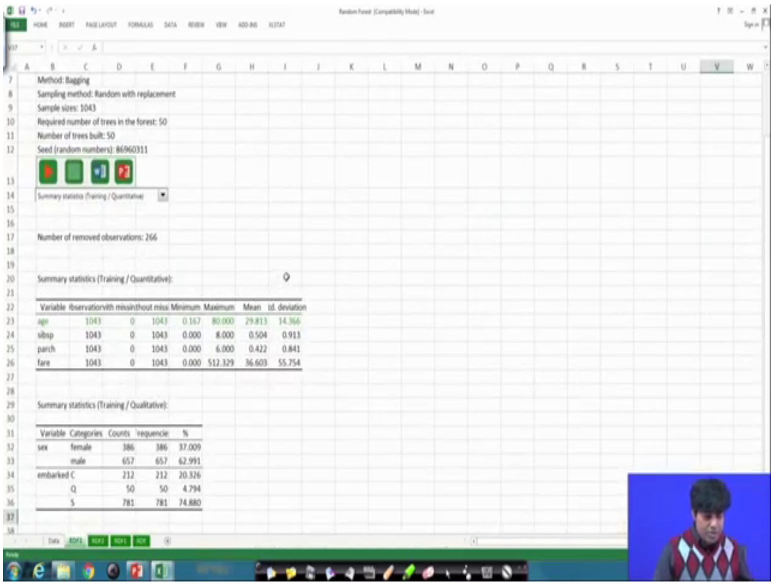
So, we have a couple of you know items here like this, and then we go to the machine learnings. And in the machine learnings we have a couple of you know tools are here, and the tool which you need right now, to forecast the particular structure is called as you know random forests. So, accordingly if you click the random forests then the particular structure will be coming like this, and then we now clean the particular you know set up first, and then we indicate the particular you know requirement. So, by the way. So, this is now the kind of you know input box, and here we like to you know classify a particular structure and do the kind of you know predictions with respect, to couple of you know variables where some are you know dependent some are you know independent.

So, in the random forests accordingly we have actually classifiers and then there are explanatory variables, and that too we have a qualitative box and we have a quantitative box. So, first hand choice is you know what is our objective, that too you know classify and that too we need actually kind of you know predictions. So, that is that is regarded as a response variable. So, in this case we like to a actually classify the p class, that is we have actually 3 types 1, 2, 3. So, in the first case response variable will be p class and then. So, we come to the quantitative variables, these are the means the quantity independent variable through which you will classify the particular you know structure, and do the kind of you know prediction.

And in this in this spreadsheet we have a couple of you know independent variables here, starting with you know age, then siblings, then the pre-class, then fares, and then this these are all actually all independent variables and that too quantitative in nature, and again come to the qualitative structure and the qualitative structures one particular variable is the survived, and then we have a sex, and then we have kind of you know envier. So, it is a kind of enough port 3 types of you know port we have recognized S C and V and accordingly. So, we really have actually observations levels and the observation levels will be the kind of you know name and names so; that means, now technically the kind of you know entry is ready so; that means, we develop a particular structure here.

So, here is our target is to classify the p class, that is the, that is actually we have a 3 here and then. So, will predict the kind of you know structures or classify the particular structure with the set up you know independent variables, that too some are you know quantitative in nature, and some are you know qualitative in nature. So, after giving the in particular you know indications, you just allow the structure to start, then this will be the p class then the quantity variables here we have e to h and against qualitative variables, this is this survived that is the a, and against the d column sex, and then the last column, the part of destiny, then finally, the observations labels that is the name clusters and then finally, put up a and continue yes.

(Refer Slide Time: 18:05)

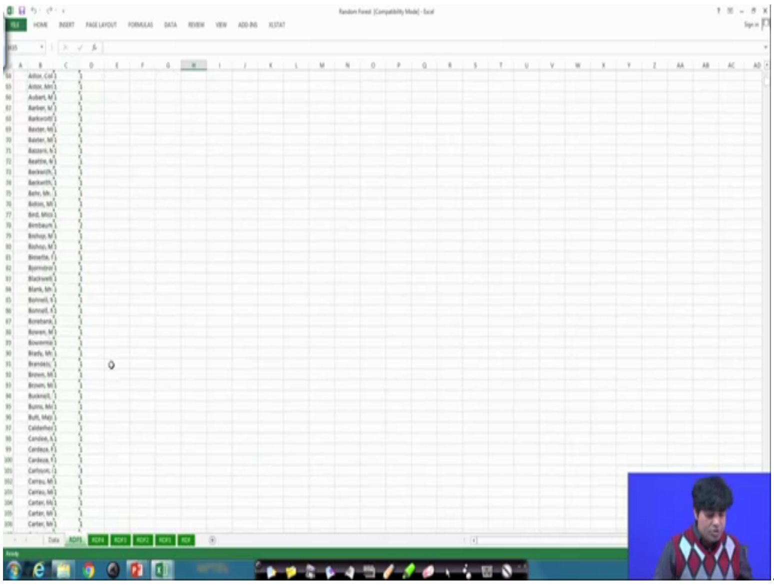


This is now ready to analyze, and this is what the kind of you know structure, through which you can do the kind of you know analysis. So, this is what the typical you know results. So now, we have a couple of variables, and the first part of this result is summary statistics. That too you know training say with respect to all these quantitative variables, and we have a couple of observation here, 1043 observations, and there is no missing items so; that means, it will give you the details, then we have a minimum of this particular you know series maximum of this particular series, mean standard deviations.

And so, means you know as usual descriptive statistic are reported to analyze the you know particular situation, and again so, then the summary statistic with respect to qualitative variables. So, this is actually sex and back. So, so these are the kind of you know structures again with respect to you know kind of you know set up and then finally, we will go for the kind of you know analysis, and then we have the observations errors. So, what will you do? We will start against to this is fine, continue something.

So, we have to give the option here. So, that you know then we will get the details. So now, and this is what the final outcome. So now, this is and this is how the final results so; that means, if you do not give the options you know the results will not actually appear.

(Refer Slide Time: 20:36)



So, initially we have given the option only for you know descriptive statistic that too for dependent variable and independent variables. Now putting all these options as per our requirement, so the final results of you know random forests is like this. So, this is what the confusion matrix. So, we have a destiny. So, 1, 3 destiny 1, 2, 3. So, with respect to 1, 2, 3, we have a 3 into 3 matrix. And then we like to see how much actually correct predictions and how much are you know the kind of you know doubtful case confusion case.

So, for a for the k you know situation; that means, classification the actual kind of you know effect and the final the predicted kind of you know effect so; that means, technically. So, we like to check 1, 2, 1, 2,2,2 and then 3,2,3. So, that is the kind of you know correct kind of you know situation. So, in this kind of you know situation so, here we have actually 100, 1043 data and accordingly. So, the 1, 2, 1 so, we have a 266 so; that means, the correct prediction is a 96 percent and then we have actually 2, 2, 2. So, that is 236.

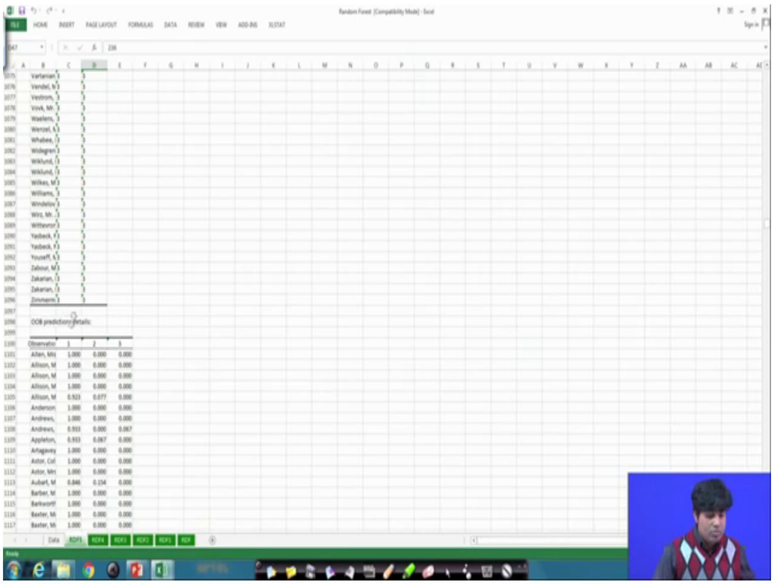
So, the correct prediction is again 90 percent, and against we have a 3, 2, 3. So, that is 485. So, it is again 96 about 96 percent, then the total prediction is actually accuracy is a 94 percent so; that means, technically. So, these are all confusion situation like you know

1,2,2. So, initially the kind of you know, set is a 1, what it is actually finally, a going to destiny 2, and similarly a the actually is 1 and then they go to destiny 3. So, likewise we have all entries, but correct prediction percentage is very high so; that means, So, the model which you have developed here to train the kind of you know structure, and to know the particular you know classification, and to find out the registration prediction structures. So, we find actually the kind the random forest give, the right accuracy having actually more than 96 percent kind of you know confident structure, through which you will do the similar you know structure kind of you know prediction.

So now, here is the kind of you know compare comparative analysis. So, the actual kind of you know fact and then the predicted effect. So, you will find see here the actual response, then the prediction response. So, 1,1,1,1 so; that means, this is what the actually the confusion matrix tells about and; that means, 266. So, this 1 to 1, 1 to 1 we have as 266 number similarly 2 to 2. So, we have a 236 numbers, and again 3 to 3 we have a 485 numbers.

So; that means, these are the correct classification. So, likewise we have actually a entered data set. So, with respect to the actual structure and the kind of you know predicted structure. So, this is 2,2,2 and likewise we have a the kind of you know 3,3,3 then; obviously, we have some kind of you know confusion kind of you know structures 1,2,2 then 1,2,3. So, 2, 2, 3 like this. So, these are all the kind of you know correct entries and accordingly.

(Refer Slide Time: 24:00)



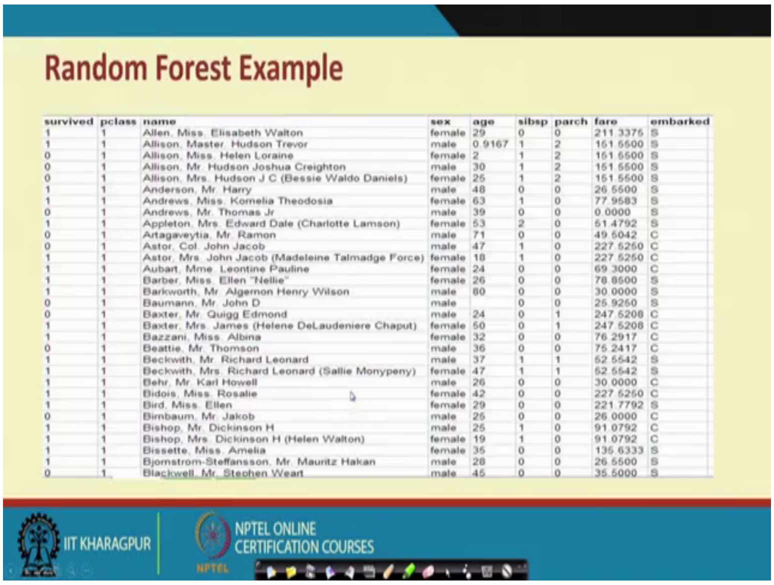
So, this is this is a more you know information about these predictions, that is actually the NM wise. So, you see here. So, these are all the 3 destiny, and if the correct destiny is 1, then by default this you know this will be 0 and this will be 0. So, likewise, altogether it should be exactly equal to 1. So, that is what the prediction structure we have set up, and here it is 1 means by default these you will be 0, since it is in 0.9 0.933. So, by default, this will be 0.067. So, as a result, this is again coming 1. So, that is how the kind of you know structure.

So, these are all for you know the, you know name wise for all the lists, and this is how the classification structures through, you know random forest. So, we have all the least a 1043 list to know the particular you know, the particular structure that is the actual fact, and the kind of you know predicted fact. So, this is again the kind of you know you know summary sheet of this entire data set, and likewise we have actually a the kind of you know structure, through which random forests can give you the, you know format through which you cannot the actual fact, and the kind of in a predicted. Fact of course, you know.

So, this is how the kind of you know a decision tree and these are the kind of you know requirement, and this is again and distinguish the kind of you know classifications so; that means, actually the fact is that you know. So, random forest give you the kind of you know classification. So, the in the we have a lots of results starting with you know descriptive statistic, but the particular you know prediction structure depends upon this kind of you know confusion matrix. So, what is the actual kind of you know you know fact and what is the kind of you know predicted fact.

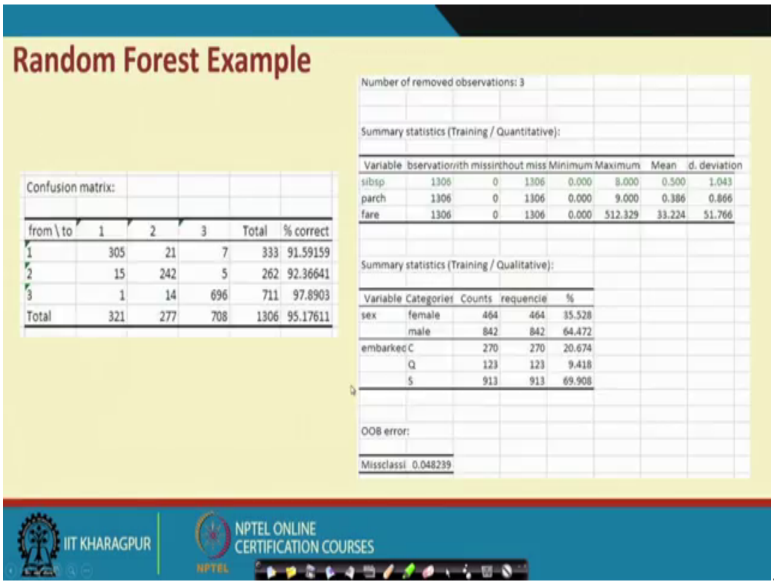
So, it is with respect to you know destiny, we are you know doing the kind of you know structure and getting the reality, how it is actually deviating from the actual kind of you know structures. So, this is how the random forest can help you lot to develop a kind of you know scenario, and then you know predict the kind of you know environment..

(Refer Slide Time: 26:18)



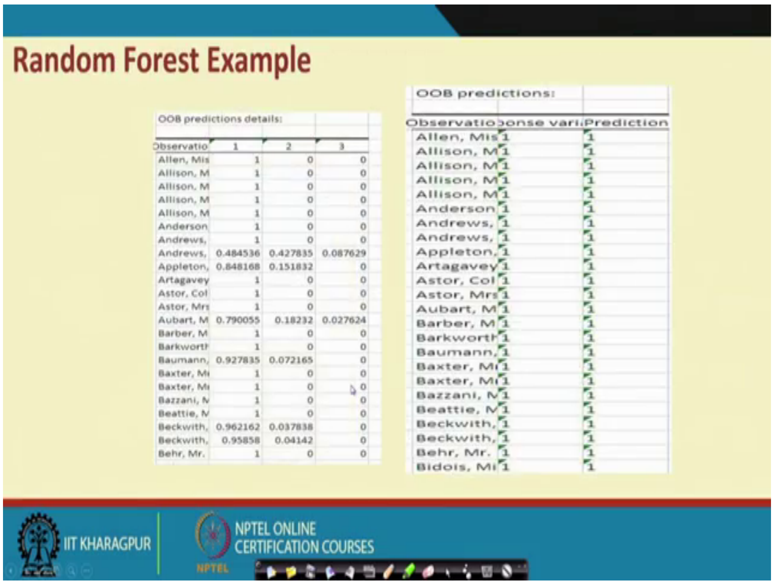
So now, coming to the kind of you know.

(Refer Slide Time: 26:20)



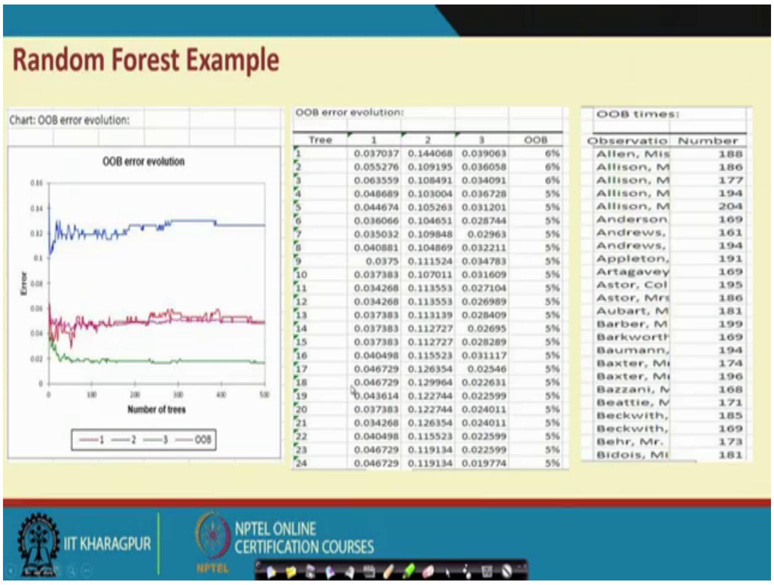
Likewise, these are all the results.

(Refer Slide Time: 26:20)



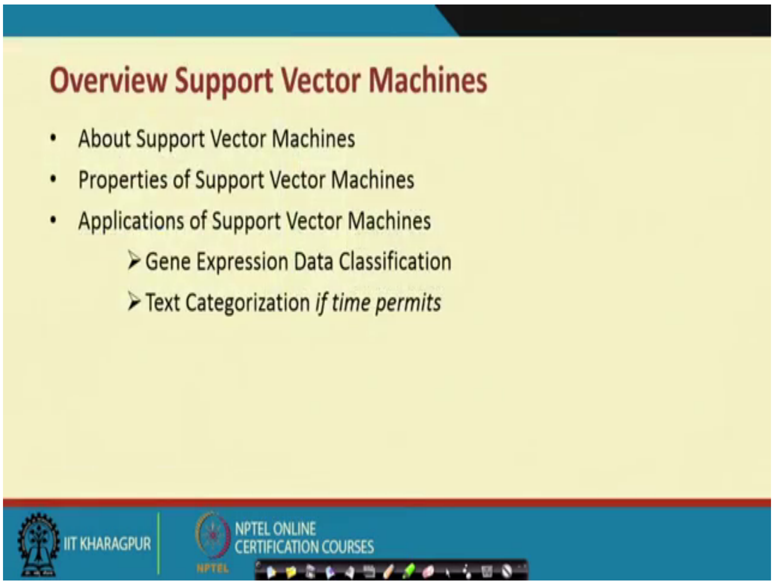
And which we have already highlighted.

(Refer Slide Time: 26:21)



And so, or the second part of this particular.

(Refer Slide Time: 26:04)

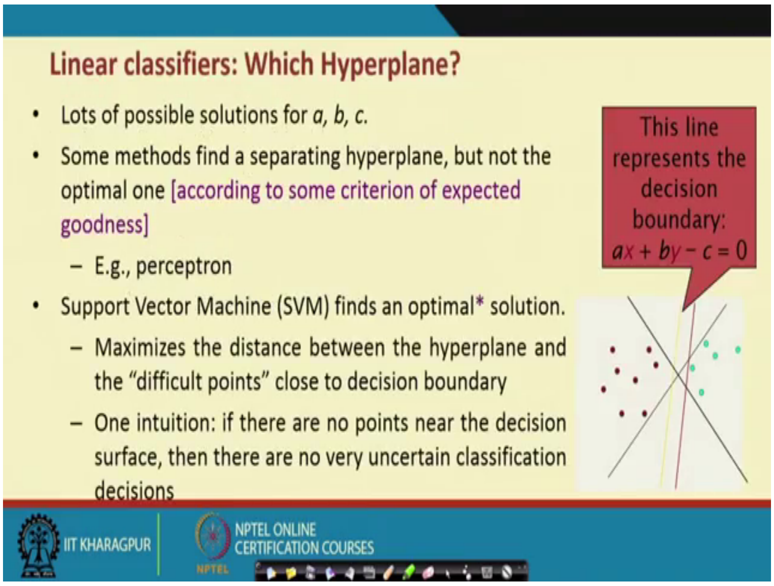


You know machine learning is the support vector machines, like you know random for the same problem. We can also classify the particular you know structures through support vector machine, and you know like say random forest it is also one of the classifiers, through which you can you know classify the particular, you know structure and do the predictions as per the business requirement.

And in the support vector machine this is actually a big component, and you know we just highlighted the particular you know structure, through which you we can connect the same problems, and then we check how the support vector machine, can give you the kind of you know classification, better classification through actually read the kind of you know business requirement or the kind of you know problem requirement.

So, we like to know a little bit about the you know support vector machines that that is s v m, and we like to highlight this you know properties about the support vector machines, the kind of you know advantages, and the kind of you know applications. In fact, we are applying this particular technique for this particular you know titanic problem..

(Refer Slide Time: 27:30)



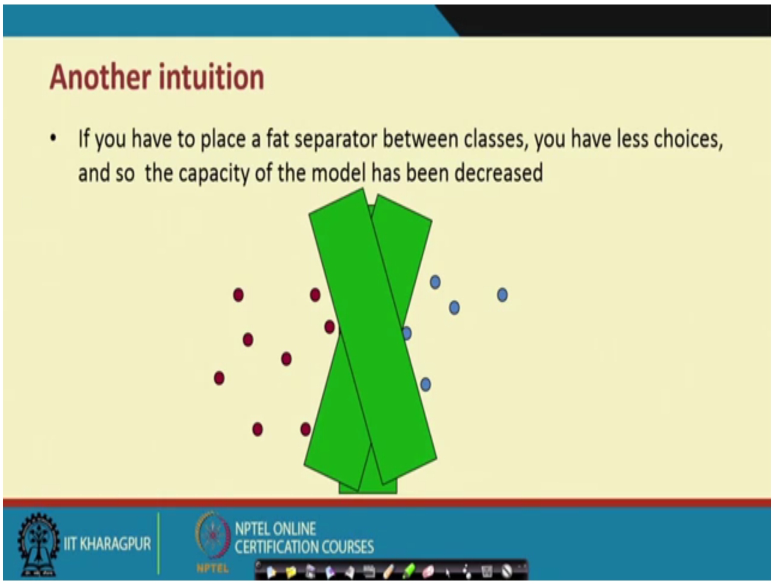
And first of all, we like to know what is actually support vector machine, what I have told you this is a kind of you know classifier. So, we can actually classify or classify the particular structure through linear kind of you know structures, and through non-linear structures. So, we have a kind of you know, but you know, kind of you know initial structure and then through the kind of you know, training and the kind of you know, algorithm. So, we can develop you know alternative kind of you know structure, through which you can you know do the better prediction, and you know kind of you know better for castings.

So, let us assume that you know this is a kind of you know lines, and represented by a x plus b y minus c. So, let us assume that you know this is a kind of you know decision boundary. And then we live to try to find out you know. Nearest kind of you know structure through which usually do the kind of you know a predictions. As per the you know particular you know problem requirement. So, support vector machines finds in a optimal solutions. Which actually maximize the distance between the hyperplane. And the difficult points close to the decision boundaries so; that means, like you know initially we have discussed about the random forest.

So, we have a kind of you know the kind of you know. Actual structure and through training and the kind of you know algorithm we develop you know alternative structure, and that that should be very close to the kind of you know actual structure so; that means, like you know whatever techniques we have already discussed, starting with you know time series technique, neural network random forest. So, here actually you know the structure which would like to finally, you know pick up and then you know apply for the prediction and forecasting. So, the since the error component should be actually drastically low, like you know we have already discussed you know mean squared error, mean absolute percentage or. So, here also same thing.

So, the particular you know structure and then we develop a new structure which the. So, the distance between the actual structure and the predicted structure should not be deviate much, then we can say that you know this particular you know classifier is a better technique through which you can do these similar kind of you know predictions, as per the a particular you know problem requirement or you know management requirement.

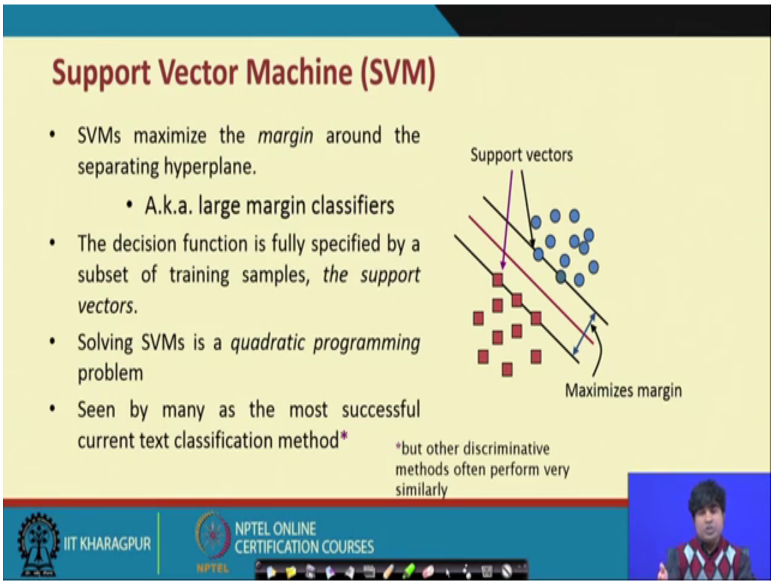
(Refer Slide Time: 29:39)



So, we have actually you know there are different ways you can classify. You know so, we have to develop the particular you know structures like you know this is the, you know kind of you know classifier. So, we have a two-different set altogether. So, the red ones and the kind of you know blue ones. So, likewise we can have another kind of you know structure through which again this can be classified so; that means, we have a this is another classifiers.

So, likewise we have actually plenty of you know classifiers, means technically against we have a kind of you know flexible structures. So, we have and we like to develop a kind of you know structure, through which you can classify the things in two different you know homogeneous group, through which will do the kind of you know similar you know prediction, like we have you know discussed the problem in the random forest. So, like you know we in the last problem we classified with respect to you know destination. So, 3 3 different destination and then the entire you know data can be classified as per the particular you know a requirement, this is also similar kind of you know structure.

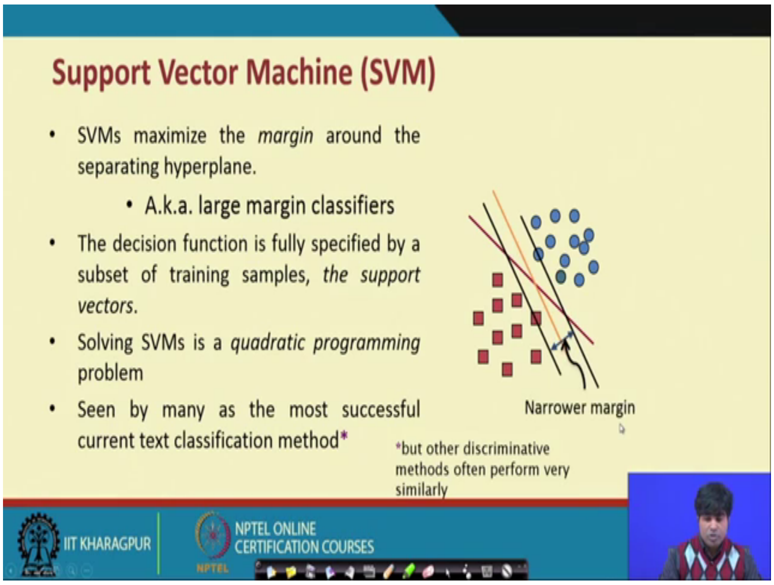
(Refer Slide Time: 30:42)



And then you know means the particular you know tool, that is you know predictive analytics tools that is the support vector machines is a kind of you know quadratic programming problems, through which you actually we really do the kind of you know classification and the develop a particular, you know structure through which you can actually you know do the prediction and forecasting.

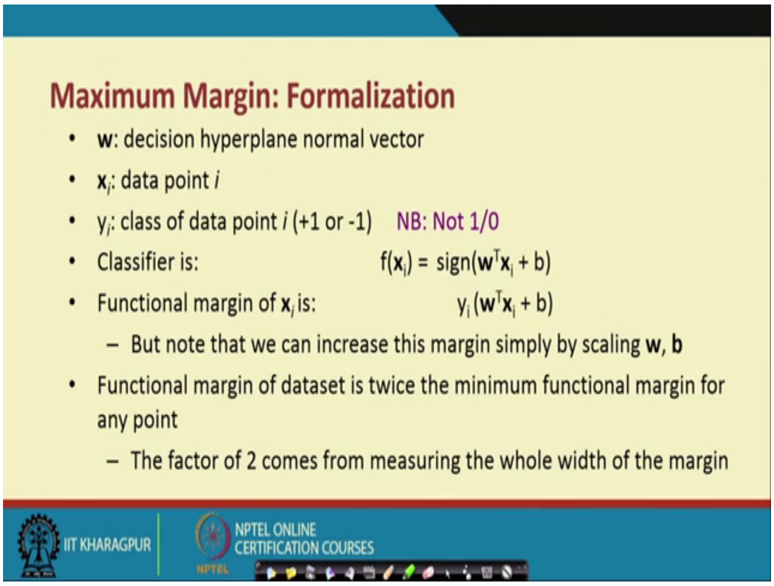
This is another way to you know do the kind of you know you know clustering. So, this is one group and this is another group so; that means, we technically get to know. So, the kind of you know structure, through which we can actually you know classify the particular you know structure. So, this is how the kind of you know actual structures, and the kind of you know derived structures, or you know new structure through which you can every times compare and you know you know check then finally, fix as per the particular you know best requirement right.

(Refer Slide Time: 31:36)



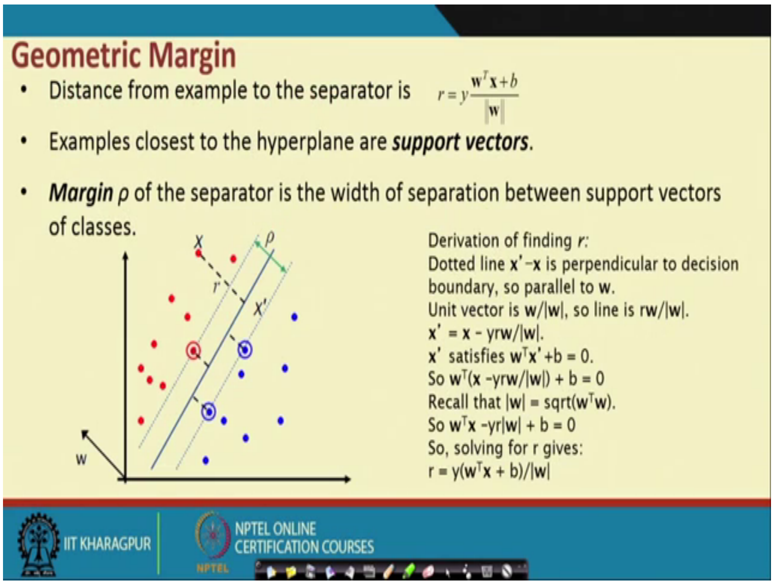
So, this is how there are different ways you can actually classify, and then we every time check and then finally, fix which is actually good for this kind of you know classification, and the kind of you know problem you know problem requirement.

(Refer Slide Time: 31:44)



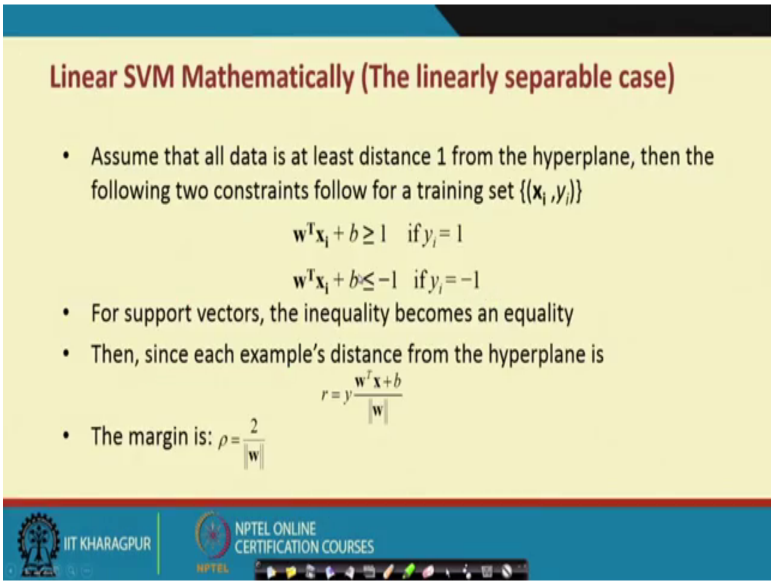
So, here the idea is you know to find out the maximum margins, particular for the particular you know process. So, this is how the simple you know functional form through which you can actually you know develop the particular structure through which you can do the kind of you know predictions, w is the decision hyper plane normal vectors, and x i is the data points, and y is the class of data points for i, that will be in between plus minus 1. So, likewise actually we like to develop the particular you know structure then finally, we pick up that structures for the prediction and you know structuring.

(Refer Slide Time: 32:25)



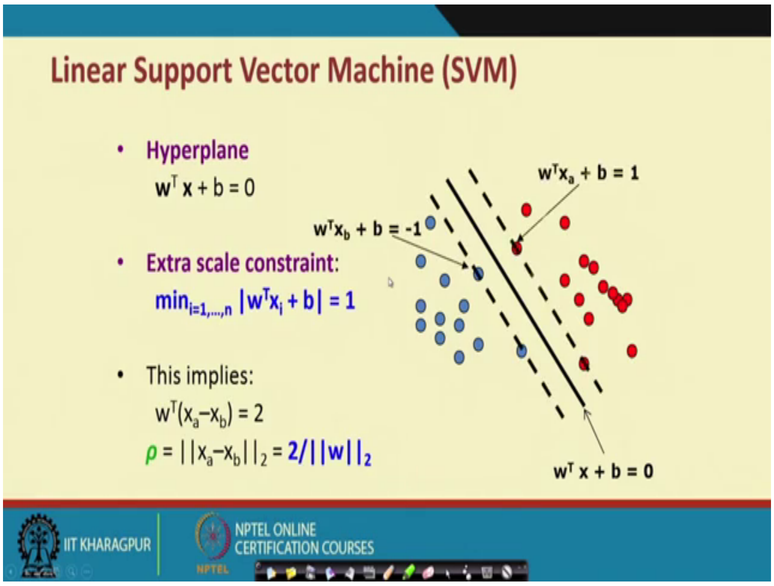
So, these are all actually the you know graphical structures, or the mathematical structure, through which you know support vector you know machines, can give you the kind of you know, flow and the kind of you know, kind of you know algorithm, through which actually you get a particular you know set up through which you can do the a better classification, and the kind of you know better predictions.

(Refer Slide Time: 32:51)



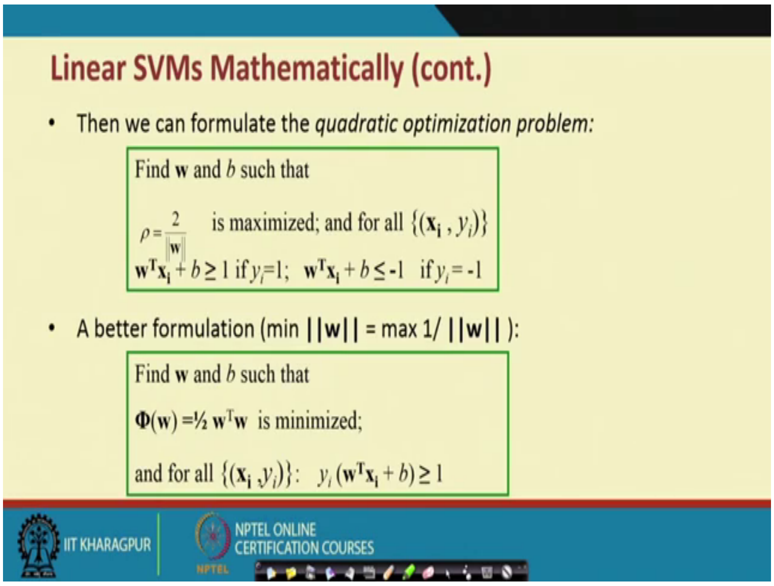
So, we can what I have mentioned already. So, we have a kind of you know linear structure and non-non-linear structure, through which support vector machine can generate a kind of you know, classification through which you can do the best kind of you know understanding the problems, and the kind of you know commodity of a kind of you know decision through each problem, can be managed you know as per the particular you know requirement. So, these are all actually linear structure and these are the mathematics behind the support vector machines.

(Refer Slide Time: 33:18)



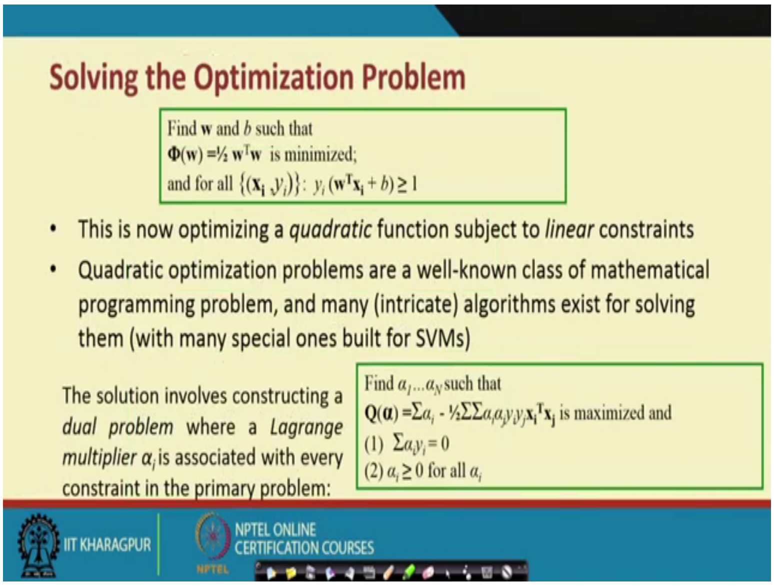
And accordingly.

(Refer Slide Time: 33:23)



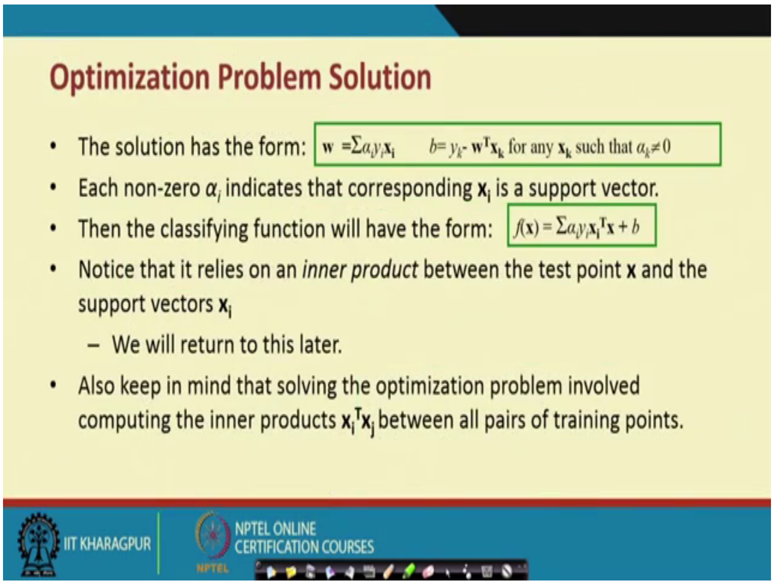
So, you know and these are all you know.

(Refer Slide Time: 33:24)



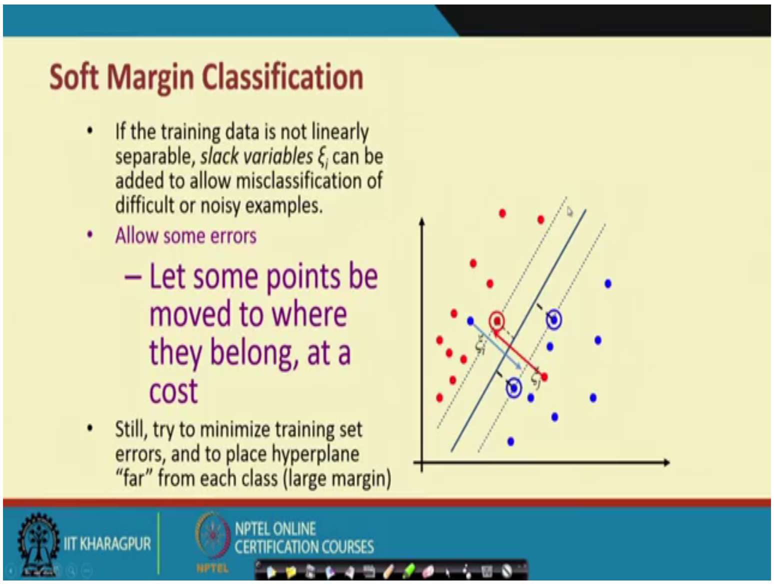
Mathematics behind.

(Refer Slide Time: 33:26)



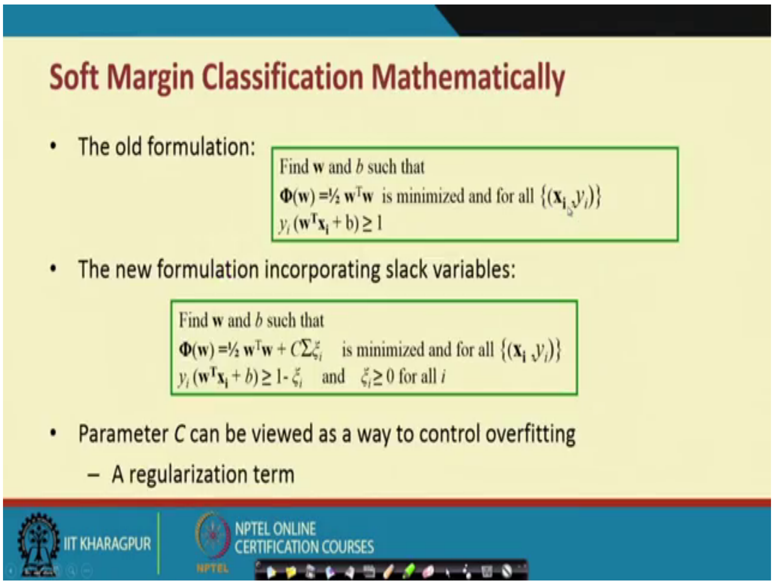
The a support vector machine.

(Refer Slide Time: 33:28)



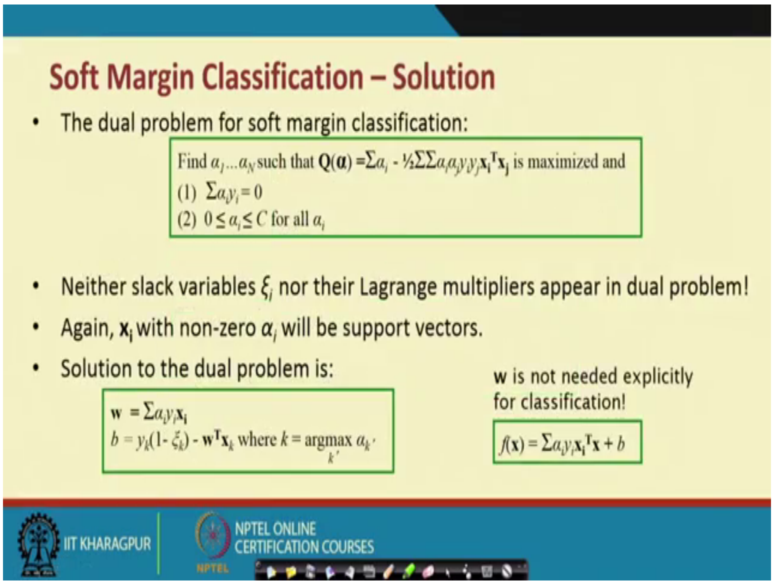
So, I am not going in details about the mathematics, but the structure of this particular, you know requirement is a say you know to find out to the kind of you know setups, and the kind of you know structure through which you can classify the particular group into, you know kind of a system through which you can, do these you know problems you know predictions and the kind of you know management requirement. So, this is more about the kind of you know.

(Refer Slide Time: 33:55)



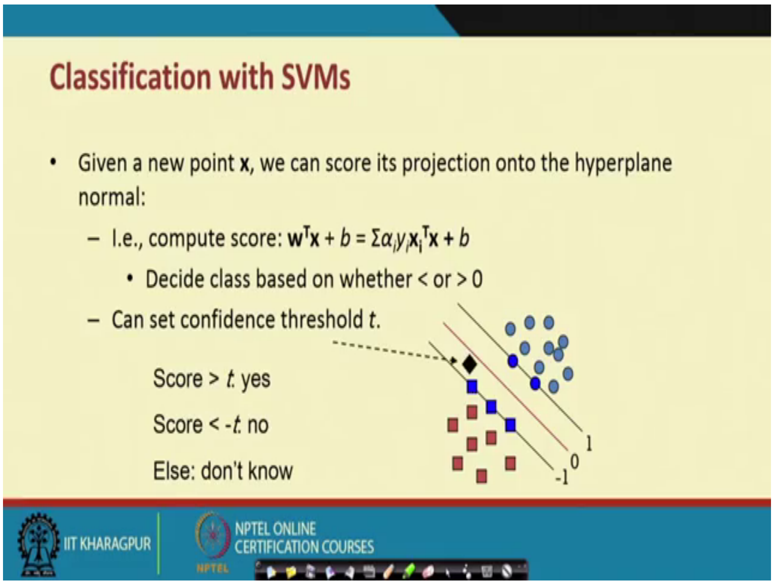
Structures and these.

(Refer Slide Time: 33:57)



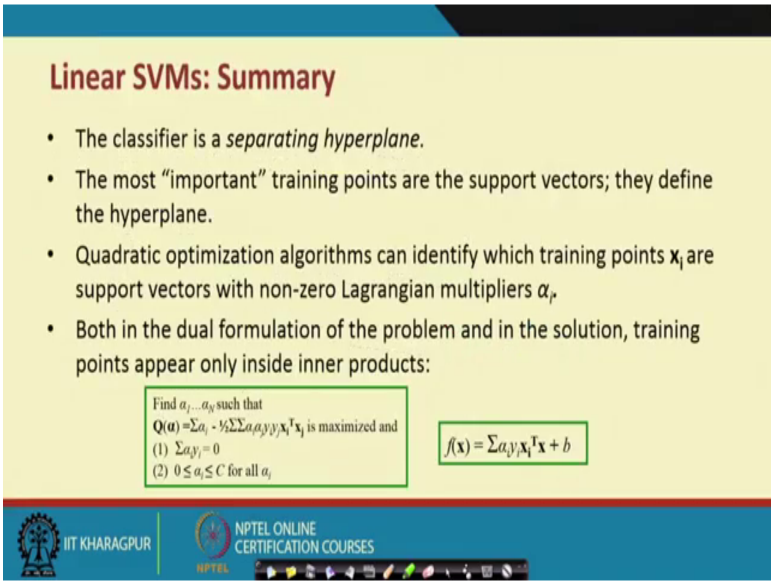
Are all mathematics.

(Refer Slide Time: 33:58)



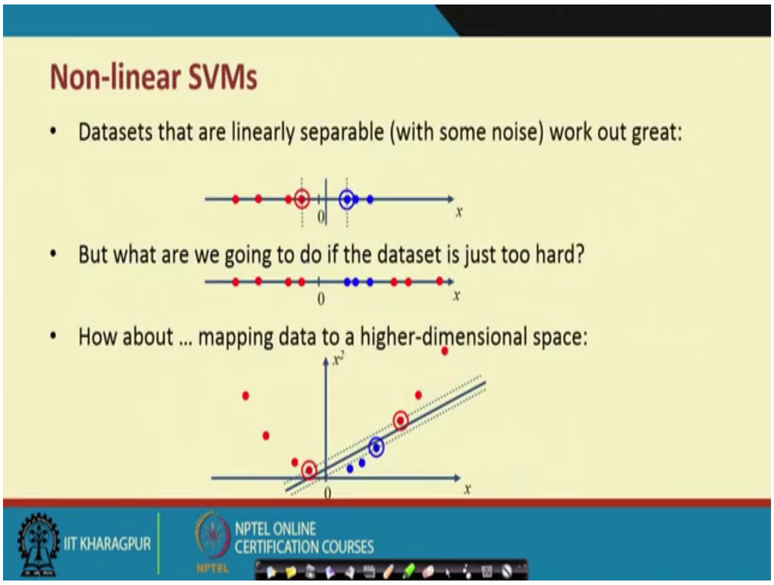
Behind classification of you know.

(Refer Slide Time: 34:00)



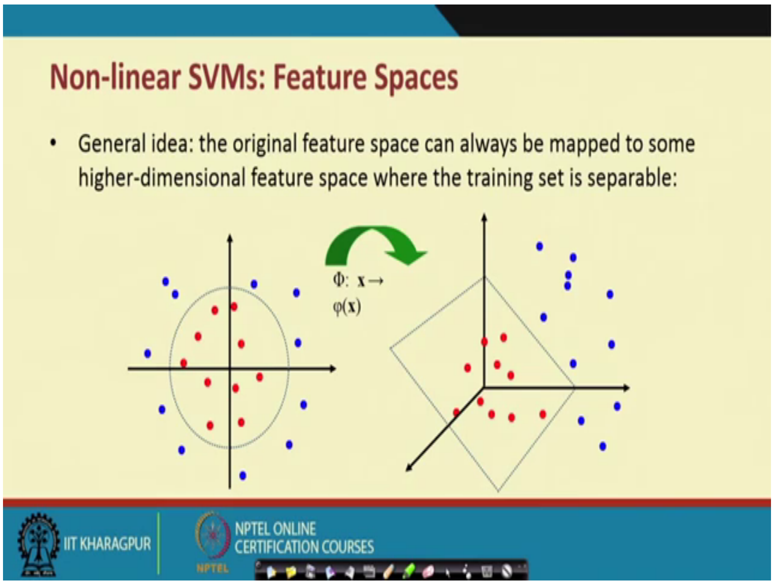
Support vector machine and this is what.

(Refer Slide Time: 34:02)



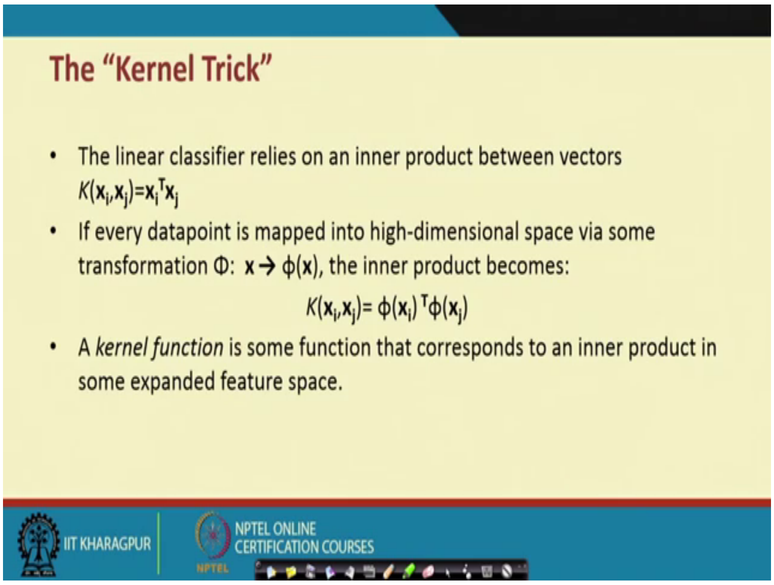
The linear and not these are all non-linear.

(Refer Slide Time: 34:05)



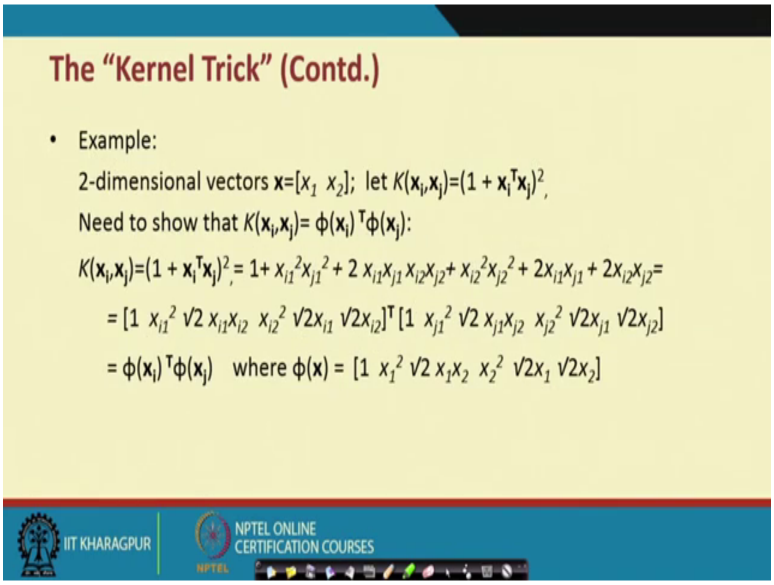
Structures which really do this similar kind of you know.

(Refer Slide Time: 34:07)



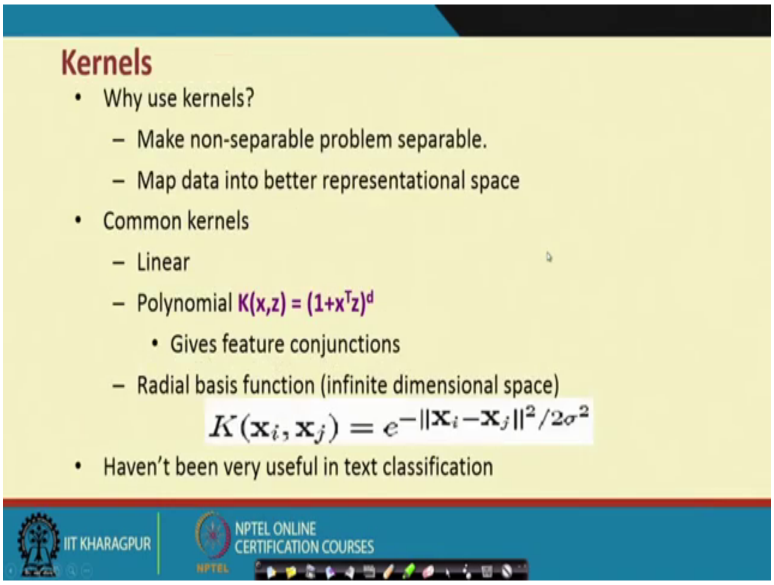
And the particular.

(Refer Slide Time: 34:10)

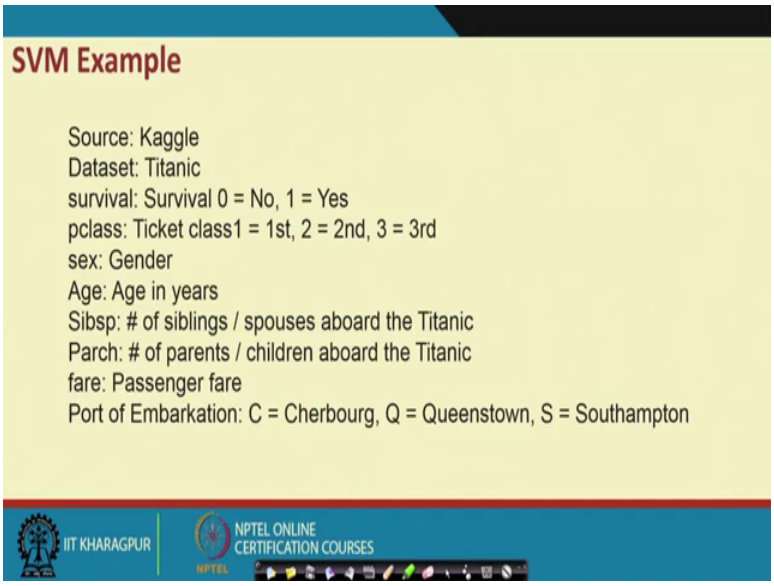


You know are follow structure is in a kernel trick, through which you can actually classify the particular you know you know structure into 2 different groups, or 3 different groups, through which you can do the similar kind of you know predictions.

(Refer Slide Time: 34:23)



(Refer Slide Time: 34:25)



So, in order to know, it is kind of you know prediction and the kind of you know forecasting. So, like random forest. So, we pick up the similar kind of you know problems, and then of course, we have discussed the kind of you know support vector math, you know machines mathematically and that too in a linear structure and non-linear structures, whatever you know mechanisms we like to follow there is no issue about it, but ultimately you like to find out a kind of you know structures, which can you know classify you know or which can you know predict the kind of you know structure through, which you will get the best outcome or you know best output as per the problem requirement.

So now, the same problems we can actually solve here, through kind of you know support vector machines then again. So, we go to the a kind of you know data set, and if she hears it is the same data set, which we have already solved through random forest and again. So, in the case of you know support vector machines. So, just we have to change the technique, because both are you know similar kind of you know classifiers and through which you can classify, and then you know on the basis of you know like you know decision tree, then you know you the predict the kind of you know situations as per the particular you know requirement..

So now again we go to the machine learning techniques. So, like you know in the last problems we have actually choose this particular random forest. So now, we have to choose actually the kind of you know support vector machines against in the support vector machines. So, our you know we will change the response variable here. So, we put the response variable survived. So, then against, we have actually here qualitative variables, and we have also set up you know you know independent variables, that is actually a quant numeric so; that means, technically.

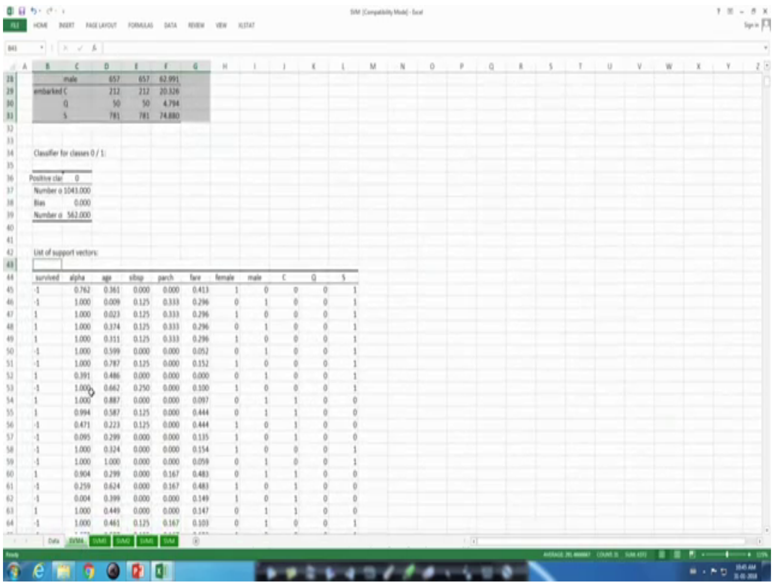
So, there are you know lots of independent variable through, which actually we have to predict the particular you know environment so; that means, the p class will be. So, this is actually no. So, here actually we put you know change the particular you know structure, in the in the case of random forest we put p class, but in the case of you know support vector machines, we take the kind of you know response variable survived and then.

So, we will go for the a kind of you know indication about the indicator independent variables and accordingly. So, the same independent variables first we put you know the numeric ones, that is the you know kind of you know of quantitatives, then against we go for the qualitative variables in the case of qualitative variables, in the case of real random forest, so, we put actually the survived, since here survive is the response variable by default pre class will be the one qualitative variables, and then against we have sex, and against we have the last row m, but. So, then we put actually. So, then by default so, you will get the output of you know random forest..

So, this is also similar kind of you know structures. So, as usual like you know random forests. So, in the support vector machines, the first in a first output with respect to these particular you know data, is the descriptive statistic you find actually compared to random forest. So, these are these outputs are you know more or less same, because these are all descriptive statistic, and this is the pool of you know you know quantitative variables, that is in quantitative independent variable, and this is the pool of qualitative independent variables, and then against we have here actually the main target response variable is the survival and accordingly. So, we will have actually here classifiers, and the bias. So, the bias is 0 here so; that means, the particular you know technique is very useful for this problem for predicting the kind of you know situations.

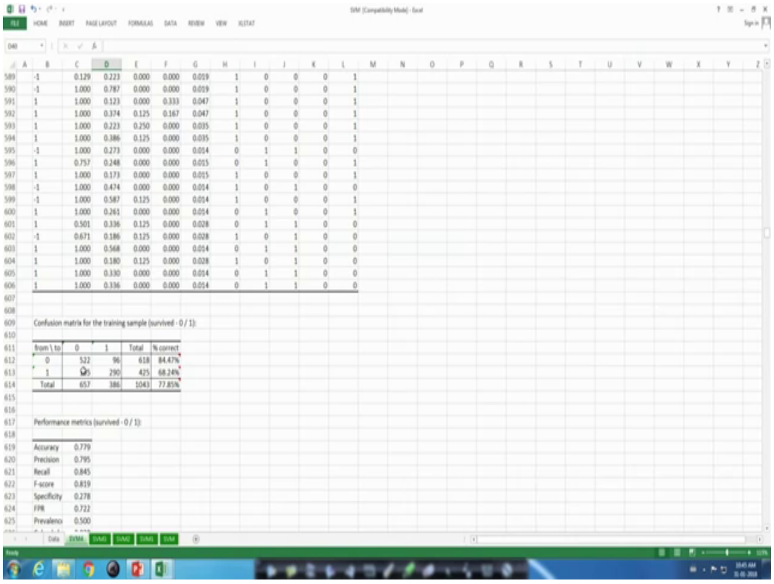
So now you see here. So, the classifier is here actually survivor. So, it is actually two kind of you know situation, if you need check the kind of you know data. So, it is actually the response is 0 1 so; that means, survive not survive.

(Refer Slide Time: 38:41)



So, accordingly. So, we have a two different classifications, and now so, this is the dependent variable response variable you know kind of you know structure, and these are all alpha is the kind of you know component that is through which you know support vector machine, give the kind of inner structure to develop the classification, and these are the you know independent variables and of course, you know the summary sheet will be prepared accordingly. So, these this is how the classification is done through support vector machines so; that means, actually we have actually plenty of results.

(Refer Slide Time: 39:11)



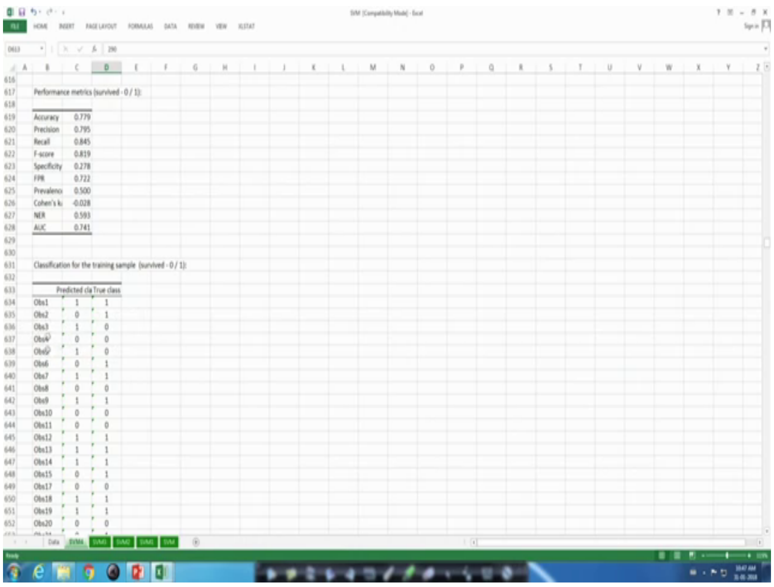
And again. So, there is a kind of you know confusion matrix. So, in the case of you know random forest. So, the confusion matrix is with respect to 3 to 3 because you know there and there are you know 3 categorical items, that is with respect to you know p class. Now here it is with respect to only 2 because we are you know putting the cost structure response or the kind of you know requirement is the survival not survivable so; that means, 0 1 of summary. So, 0, 1, 0. So, by default we have 2 into 2 matrix and then again. So, 0, 1, 0, 0 to 0. So, this is the correct kind of you know response, and 1 to 1 this is the correct kind of a response, and compared to random forest with respect to response variable p class.

So now in this case the kind of you know correct prediction is the 84 percent, and you know in total it is coming around 78 percent so; that means, So, in the kind of you know p class classification means in the response in the case of you know response p class, then the classification structure gives you know better kind of finished, you know prediction environment through which will you do the prediction, and the kind of analyze the particular you know problem, but here with respect to change of the response variables to survival. So, here the correct prediction is not actually.

So, good compared to the, you know response variables, the kind of you know p class, but by the way. So, if you come if you compare the results. So, it gives you know better, because this is you cannot you know actually you know just ignore, because it is the, it is because of you know change of response variables. So, the if you actually use the same response variable in the case of you know random forests, most probably the result will be coming like this.

So; obviously, so, in this case. So, the performance matrix will be also like this, this is the accuracy position 78 percent, and likewise the EPS course all these are you know different, you know model indicators or you know accuracy indicators through, which really do the kind of you know we can analyze the problem..

(Refer Slide Time: 41:24).



So now, here the kind of you know comparative analysis. So, the kind of you know observation wise. So, the predictive structure and the true structure, then you can also this is the actual kind of you know, and these are all actual kind of you know structure, and these are all predicted structure. So now, you can check actually, this is predicted then the actual so; that means, there is no error here, but here there is error. So, this is the 0 and 1 so; that means, these are all changing situations, like you know you have a here you know confusion matrix.

So, either the situation will be correct or something some change is happening so; that means, if 0 is the actual and the prediction is 0. So, this would be under this particular you know entry if it is 0, but actual happening is if the predicted entry is happening 1. So, in that case it is actually 96, similarly 1 is the actual fact and the predator structure 0 is coming 135, and then against 1, 1 so; that means, the actually is 1 and the predicted is 1.

So, that is the number is 290. So now, if we add all these figures. So, it is coming 1043 so; that means, the confusion matrix gives the correct structures through, which you can understand the particular in our requirement, and understand the kind of you know structure through actually do the kind of you know predictions as per the particular you know business requirement, or the kind of you know problem requirement. So, technically, we have already discussed 2 different you know machine learning techniques here. So, these are all you know summary of these results, what we have already discussed with you know support vector machine, this is what the original problem.

And this is what the summary sheets, and these are all again the kind of you know basics about the qualitative descriptive statistic, and these are all intermediates output, and this is what again training state, training structures, validation structures, and finally, the outcome the measure outcome of this particular no support vector machine is the kind of you know confusion matrix. So, that is what the actual validation is happening like the case of you know you know random forest.

So; that means, technically what I like to say that you know in the machine learnings whether it is artificial neural networks, or random forest, or support vector machines. So, these are all you know you know very interesting techniques, through which you will do the predictions, typically when we have actually big pool of you know data, or you know big set of you know variables, and by the way we have no clear-cut idea about the particular you know structure.

But the neural network and the random forest and the kind of you know support vector machines, by default over the you know training and the kind of you know test you know flexibility. So, we will get some kind of you know best structures, or you know beautiful structure through which actually you can classify the data understand the data, in a kind of you know much better framework, and then you can predict as per the particularly no problem requirement and the kind of you know mismanagement requirements.

So, once you actually pick up a particular you know technique, best technique as per the particular you know problems, compared to you know the data size, and the kind of you know number of variables and; obviously, So, the decision making will be very effective as per the particular you know problem requirement. So, with this we will stop here.

Thank you very much have a nice day.