Hello guys. So we are going to continue the discussion with respect to a new machine learning algorithm, which is called as nave bias algorithm. Now, nave bias algorithm is specifically used to solve classification problem both binary and multi-class classification. In order to understand nave bias right, you really need to have some basic idea about probability. Okay, so super important. There is something, some concepts in probability that you really need to understand. Okay. And after understanding this concepts we will also try to derive an important theorem which is called as Bayes theorem. So based on this Bayes theorem this entire nave bias algorithm machine learning algorithm works okay. Now let me just go through this probability. Like what all topics we specifically need to revise okay. So uh, in probability right. Uh, with respect to various events. Right. There are two types of events. Like there is one something called as independent events. And the other one is something called as dependent events. First of all, let me just talk about an example of independent events. Let's say I am rolling a specific dice. Now you know what are the outcome of a dice? It is nothing but one, two, three, four, five and six, right? So these are my specific outcomes with respect to a dice whenever I roll. Now, whenever you try to find out the probability, let's say that I want to find out the probability that the roll should be one right. Or the outcome should be one. Then what will be the output over here? First of all, you will divide by total number of outcomes. So over here you have 123456. And then probability of one will be that. How many times one can come in a in a in a single roll. Only one time. Right. That is the maximum probability that you can get. So the overall probability of one is somewhere around one by six. And even similarly, if I go ahead and calculate probability of two I will be getting one by six. The probability of three will be nothing but one by six. Right. So this is with respect to an independent event. Right. So here why do we say this as independent event. Because one outcome is not hampering or not changing the probability of the other outcome. Right. So probability of one is not changing the probability of two. The values are same every time. When I am rolling a dice I will be almost getting the similar kind of probability right one by six, one by six, one by six. But if I talk about the second type of events, which is called as dependent events. Now in the case of dependent events, let's say what does a dependent event look like? Okay, let's consider that I have a bag of marbles okay. Let's say there are three orange marbles and there are two yellow marbles. Okay. Now my question is what is the probability? What is the probability? Okay off. Removing a white marble. White marble and then a yellow marble. Then a yellow marble. I'll not say white marble because I have taken an orange marble. So let's go ahead and right here it has orange marble. Okay. So my simple scenario is that what is the probability of removing first first event that is happening is removing a orange variable. And then you are removing an yellow marble. So in this scenario let's say first of all I remove a yellow I remove a orange marble from this bag. So what is the probability of orange. Very simple. The probability of orange is nothing but three by five. The total number of orange marbles are three. And uh, over here I have three marbles and total number is divided by five, which is my total number of marbles. So this is my first event. Okay. This is my first event. The second event. Now, after this, I definitely need to remove the yellow marble from this bag. So if I say three by five one once this event occurs, right? How many number of marbles I have remaining in this bag? Right. I have somewhere around four marbles. See? One, two. Right. Two red marbles are there. And along with this, two yellow marbles are there. Right. Orange marble has been removed. So in this event the orange marble has removed. Marble has been removed. Okay. Now, in the next event, I need to probably remove the yellow marble. Okay, so in the first event, see what has happened. I have removed the orange marble. First of all, we need to find out what is the probability of removing an orange marble and then a yellow marble. So this events are basically called as dependent event because one event is happening and the first event will definitely change the probability of the other event. Right. How it is changing. See the count of the marbles? They are only four marbles right now. If we go and see what is the probability of the yellow marble after this particular event, then you will be able to find it will be nothing but two by four, which is nothing but one by two, right? So this is the outcome of the second event. This is the outcome of the second event. Right. And this event has already been impacted by the first, uh, first event. Why? Because here initially there were five marbles, but here it became as four marbles because I removed one. In the case of independent events, my outcome will be always same for any kind of events, right? In in rolling a dice. This outcome, it will always be one, two, three, four, five, six for any kind of events. Right. So here we have specifically removed the yellow marble. Remove the yellow marble. Now if I try to combine both these events I can use a notation. Right. What is the probability of taking out the orange marble first. And then the probability of taking out the yellow marble. Right. Very very very simple. Then over here I can probably denote this entire equation by probability of orange multiplied by probability of yellow. Given orange event has taken place. Given orange features have or orange, I will not say feature orange marble has been taken out. Right. This is what it basically indicates here. What I'm doing. What is the probability of orange. And then the yellow marble to take place one after the other dependent events. Right. So here I will basically be writing the probability of orange multiplied by probability of yellow given orange has taken place. And if I really want to find out this value then what exactly it is here I can write this as also probability of orange given sorry, probability of yellow given orange has already taken place. Right? And this is my second event. This is my first event okay. This is my first event. This is my second event. And I hope you are able to understand now right. This scenario is nothing, but you're taking out the yellow marble. Given the orange marble has already been taken out or the first event has already been taken place. So what do we do? We basically combine this. And this is this equation. This this side equation is basically called as conditional probability. Why do we say it as conditional probability. It is very important to understand because here I'm saying probability of yellow marble to take out the yellow marble given. Taking out the orange marble has already been done right. Or I can also say this that probability of a event considering a specific event has already occurred. That is nothing but conditional probability. So this entire thing is basically called as conditional probability. So I hope you are able to understand. Now if I combine this equation, what will be probability of orange will be three by five multiplied by one by two, which is my probability of y given O has already occurred. So here I'm going to write it as three by ten. So this will be my probability of O and y. Again I'm telling you O and y basically means I'm taking out the I'm event. A event O is occurring first and then event Y is occurring after that. Okay. Now in a generic way, if I really write need to write this equation, it is nothing but probability of A and B is nothing but probability of A multiplied by probability of b given A has already taken place. So this is my equation in this case of. Dependent event. Okay. And this is super important because this is what we are going to use for deriving our Bayes theorem. Okay, super super important. Now let's talk about Bayes theorem. As I said, first of all we need to understand probability. Then we'll need to understand Bayes theorem. Because based on this Bayes theorem, this entire Naive Bayes algorithm works. Okay. So what I'm going to do over here, first of all, let's go and write the Bayes theorem. And what exactly Bayes theorem is okay. Now just imagine, guys over here, okay? You know that if I take this equation. So if I probably take this equation probability of a and b okay I'm going to derive the Bayes theorem. Now is nothing but probability of b and a right. We can write like this. Also probability of event A occurring and then b occurring instead. We can also be. This will also be equal to probability of b and a. Also. Right. And in my previous videos also I have discussed about Bayes Theorem. Now how do I expand this equation? In order to expand this equation, I will write probability of a multiplied by probability of b given A has already taken place. Right? And trust me guys, if you do the calculation with respect to this also right, you will be getting the same output for both this kind of events that is occurring now. Similarly, on the right hand side, I can also write this as probability of b multiplied by probability of a given b. Right. Both these equations are almost same right? Over here you can clearly see right what I'm actually writing. Right. The same equation is basically coming from this. Right. The same concept I've used over here. Now let me do like this. Let me take this equation and write somewhere like this. Probability of A by b is nothing but probability of A multiplied by probability of b given a divided by probability of b, right. So this entire equation. So what I did over here I just kept probability of a by b at one side. And I divided this in below. So this equation that you are probably getting this is basically called as Bayes theorem a super important theorem. And now we will try to understand how this equation is basically used in a machine learning problem statement okay. So a very simple way to derive the Bayes theorem okay. The Bayes theorem says that the probability of A or b is equal to probability of A multiplied by probability of B given a divided by probability of b. Now what is probability of A given b that is nothing but probability of event A? I want to calculate probability of event A given B has occurred. Okay B given B has occurred. Right. So this is probability of event A. Similarly if I go ahead with the next one what is probability of a. Here I can basically write. This is nothing but probability of event A. Okay of event A, and this equation will also be used to solve machine learning problem statement. Then what is probability of B? It is nothing but probability of probability of event B right. And finally you have probability of B given A. And obviously you know what this is. Probability of event B given A has occurred. Given A has occurred. This is quite amazing. Now you have understood each and every components inside this Bayes theorem. Now I will again copy this entire equation and make you understand how this thing will now be used in with respect to machine learning. Okay. Super important okay. Now in machine learning problem statement guys, you have lot of independent features. Let's say my F1, f2, F3, F4, F5, F6 are your independent feature right? And let's consider I will write instead of writing f1, f2 f3. Let me go ahead and write. Let's say that I have three independent feature X1X2X3 and this y is my dependent feature. So I may have some values over here. Here I may have output like yes or no okay. Yes or no. So here also I will be having yes here also I will be having no. So like this kind of independent. So this are my independent features right. So this are my independent features. And this is basically my dependent feature. Right. The output feature. Now I will try to solve this entire prediction and all by using this Bayes theorem okay. This is what is a Bayes theorem. Now we need to see the algorithm that specifically use this Bayes theorem is called as Naive Bayes algorithm. Right. And it is used for binary and multi-class classification. Now how we will be using this equation to solve it okay. Now I will write the same representation of this equation. Here I will say see what we need to predict. We need to predict y right? So in short can I write this equation as probability of y given my features x one comma x two comma x three. Right I can write like this. The same thing. It is right probability of a given b. So probability of A is nothing but probability of y. I need to find out what is the probability of y given X1X2X3. Then I can probably go ahead and write this equation as probability of y. Okay. See this probability of y multiplied by multiplied by probability of x one comma x two comma x three. Given y okay. Given y divided by probability of x one comma x two comma x three. The same equation we have written in this way, and what this equation is saying that we need to find out y given X1X2X3. Okay. So this equation is all about naive bias machine learning algorithm. So I will just copy and paste it over here because we will further use this formula for a very important purpose okay. But this is the entire equation. Now if I expand this equation how this equation will now look like. See I will write like this only. So here I have probability of y multiplied by probability of x one given y multiplied by. See. When we expand this right, we have to expand with each and everything. So here then I have probability of x two given y and then probability of x three given y. See how many number of independent features I have? Not that many number of times I need to expand. Okay, now what I will do I will copy this. I'll divide this by probability of x one comma x two. Or instead of writing like this. Also I can write like this. See probability of x one multiplied by probability of x two. Probability of x three. Understand the mathematical intuition guys. Because I will solve a problem in front of you. I'll take a data set and solve this. What is probability of x one given y? What is the probability of x two given y? I will solve it with the help of example, but understand the equation. Okay, now. With respect to this particular data set. Okay. If I take this data set okay. What are the two important things that we need to find with respect to the output? Right. The two most important thing that we need to find out is that what is the probability my outcome of y is either yes or no. So I will say what is the probability of yes given x one comma x two comma x three. This is the first thing that we need to find out. So this one I can basically write it as probability of yes multiplied by probability of x one given yes multiplied by probability of x two given yes multiplied by probability of x three given yes. So this three things I need to find out from this equation. And this will be divided by what probability of x one. Probability of x two divided by probability of x three. So this two things we need to find out. See if my outcome is yes or no. I need to find out what is the probability with respect to yes. And what is the probability with respect to no. If my probability with respect to x is 0.60 and probability with respect to no is 0.40, again, we know that the maximum probability is saying yes. So my outcome will be basically yes. Right. So similarly we need to find out what is the probability with respect to yes. And what is the probability with respect to no. So here I will go ahead and write probability with respect to no. Also I will try to find out given X1X2X3. Now based on the above equation how my this equation will change. Very simple. I will write probability of no multiplied by probability of x one. Given no multiplied by probability of x two. Given no multiplied by probability of x three. Given no right. So once we divide this now, this time we divide this by probability of x one, probability of x two and multiplied by probability of x three. See I'm deriving right now. Still I'm I'm saying this. Now observe one thing over here in the denominator. Both these are constant. Because this is also I'm going to get a constant value. And here also for this denominator I will get a constant value right. So what we can do we can cut this equation. We can reject this or we can remove this entire denominator. Now if I really want to calculate what is the probability of yes given X1X2X3, I will just use this equation to find out the output. Now let's say if I get the probability as 0.60, I get the probability as 0.40. Now you know that for any new test data, and this is for any new test data. So any new test data if I'm trying if I'm getting probability of yes is 0.60. Probability of no is 0.640. Then my output will be this one. That is nothing but yes. Right now. This was almost all about the name bias machine learning algorithm. This is how it works. It uses this Bayes theorem, the simple Bayes theorem, to probably calculate everything. Right. This is my simple Bayes theorem. Right. And the same Bayes theorem you are trying to find out with respect to yes and no, and try to find out the output. Now let's go ahead and solve a specific problem where we will be considering a data set, and we'll try to solve it. The data set we'll be talking about will be an amazing data set. We will take up that okay. And obviously we have used that data set even in Decision Tree. And we'll try to manually find out what is the probability of yes or no with respect to any new data point. So let's go ahead and let's talk about that. So guys now I think we have discussed about Bayes theorem. We have understood how naive Bayes algorithm works. In short it calculates the probabilities like yes or no if it is a multi-class classification based on the input features that we are getting. And we probably got this entire equation. Also. Now what I'm going to do is that I've taken this amazing example which we have also seen in this particular decision tree. So this is the data set. We will try to find out. I'll do the manual calculation and show you what all probabilities we will be getting. Okay. Now let's go ahead and probably split this particular data set. And when I say split the data set not using decision tree we'll be using naive bias. But with respect to every feature we'll try to find out something okay. And then whenever we get a new data point we'll try to predict it. Now first of all, the first feature that I'm going to probably take is outlook. Okay. So here you can probably see outlook is there. So here let me go ahead and write outlook okay. Now based on this outlook feature what all things we really need to find out okay. So in outlook what all possible options you have. See sunny you have overcast you have and rain you have. So I'll write the three categories over here sunny, overcast and rain. Why? Because I really need to find out the probabilities of each and every thing right now with respect to sunny, how many yes outputs are there? How many? No output are there. And based on this two values. Right. I also need to calculate two important things. Okay. That is what is the probability of this sunny event. This sunny event given. Yes. And what is the probability of this probability of this sunny event given. No. Right. So these two values will be using it. Because at the end of the day we are going to use this kind of equations in our this particular formula right when we are trying to calculate okay. So uh when I say X1X2X3 over here. Right. These are my X1X2X3. These are my X1X2X3X4. Like that. So in this case what we are going to do is that we are going to solve this classification problem based on this input features outlook temperature humidity and wind. Now let's go ahead and find out what is the probability. See how many times like sunny whenever sunny out whenever sunny input is there? How many times yes are there? So if you probably go and see. With respect to sunny, there are around one year here and one more may be here, right? So two. Yes. Are there one. Yes. This side. One. Yes. This side. So there are total number of. Yes. In the case of sunny is nothing but two. Now similarly with the case of overcast how many. Yes and nos are there. Just go ahead and calculate. You'll be able to get four. And with respect to rain they will be three. Right. And in case of sunny again no. How many nodes are there. Three. Just go and count it. And we have done this in Decision tree by using the same example. Then in case of overcast it will be zero. In the case of rain it will be two. Now in order to find out this important thing, what is the probability of sunny given? Yes. Right. What is the probability of sunny given. Yes. So total number of yes. If you try to find out count over here 123456789. Right. So total number of yes are nine. And when I say probability of event probability of sunny event is how many times two given. Yes. Right. So here I'll be writing two by nine. Right. Similarly if I go ahead and calculate with respect to overcast it is four by nine. So I have this four number and divided by nine. And this will be three by nine. All these values are there now. Similarly we'll go ahead and calculate what is the probability of event given. No. Now you are going to focus on this divided by nine. So in this case I will go ahead and write. This will be three by nine, zero by nine and sunny not nine. It will be five right. The total number of nos are how much 12345. So total number of nodes are five. So here I'm going to write three by five. And then here you will be writing zero by five. And here you'll be writing two by five. Right. That many number of nodes are there. Now similarly for every feature will. Go and see what is the probabilities. Right. And then we will try to find out what is the output. And we'll use the same formula. Now this side I'm going to probably go with temperature. So let me go ahead and see with respect to the next feature that is temperature. Now with respect to temperature how many different values are there. Obviously you know over here one is hot, one is mild. See hot mild and cool. One is cool right? Now if I see with respect to outputs, I will be getting yes or no. And what is the probability of the event given? Yes. And what is the probability of event given? No. If I had more classes, I would have right probability of event given the third category. Fourth category. Fifth category. Right. Now, with respect to hot, how many number of yeses are there? Just try to find out how many number here I can see one and the other one I can probably see here two. Yes okay. So I will go ahead and write this value two. Yes. Four. In the case of mild, when the when the value is mild inside the temperature for yes I am able to get when it is cool I will be able to get one. Yes okay. Now similarly with respect to no sorry not one. So two four and with respect to cool. Also if I see right there are three one yes two years and three years okay. So just try to calculate it guys. Just try to find out how many numbers are there in case of no it will be 221 okay. So let's do the calculation. Uh automatically you'll be able to understand how many yes and nos are there. Now what is the probability with respect to this event that is hot given. Yes. So obviously you have you have to. So here I will be writing two by nine. Here I will be writing four by nine. Here I'll be writing three by nine. Similarly the probability of event given no. So I have to consider this. So it will be nothing. But when given no basically means what given no basically means uh, what is the probability of hot given? No. Like like that. So if I probably see this will be nothing. But again two by nine, two by nine and this will be one by nine. Okay. So these are my values that I am able to get with respect to all the values over here. And sorry this will not be nine guys. Again, I'm forgetting it because the total number of nodes are five right. So total number of nodes. So probability of event. When I say probability of hot given total number of nodes okay. So here I will go ahead and write probability of event. Uh that is hot hot event to take. No. So over here I will write two by five four by five and three by five. Right. So total number of nodes in this are five. Now let's go and take only this two uh features. I will not take the other two features. Let's consider only this two features just to give you an idea. Now finally the output feature is play right. And play has values either yes or no. Okay. So here also we'll go ahead and calculate the probability. So total number of yes and no are nothing but nine. And five. So if I really want to find out the probability of yes right. It will be nine by 14. Right. Total number of events are 14. So here nine by 14. Total number of yes and nos are 14. So probability of yes will be nine by 14. And probability of no will be five by 14. Now we have calculated almost everything. Right now our main aim is that whenever we get a new test data and let's say the test data in case of outlook is sunny and in case of temperature is hot, we really need to predict what is the output. Okay. So for this feature we have taken the first feature as sunny and the second feature as hot. We need to predict what is the output. Now how do I write this? I will write in a simple equation. What is the probability of yes? I need to calculate right given I have sunny comma, hot x one and x two feature right when I have this two features. So how do I write this? It will be probability of yes right. Multiplied by probability of sunny given. Yes. Multiplied by probability of hot given. Yes. Right. Very simple. I don't have to take the denominator because I told you. Right. With respect to yes and no the denominator is a constant value. So this is what is my values right. This is my entire equation to calculate according to the name bias. Right. But if I really want to go ahead and write the denominator part, this will be probability of sunny multiplied by probability of hot. And you know that this will be a constant value for yes and no both right. So I will remove this okay. Now let's go ahead and see what is the probability of yes. Probability of yes is nothing but nine by 14. Right. So now I will equate it. See nine by 14 multiplied by what is the probability of sunny given. Yes. What is the probability of sunny given. Yes. Two by nine. Right. I can see over here. See this is nothing but two by nine. Probability of sunny given. Yes. Right. So this. In short I can also write. This is a probability of sunny given. Yes. So it is nothing but two by nine. And the last term is what is the probability of hot given. Yes. Hot given. Yes two by nine again. So here I will go ahead and calculate it two by nine. So in short if I probably calculate this if I cut all this things like this. Right. So uh over here you can basically see that I am going to get uh, so nine and nine not deducted. And I have two ones are two, seven, seven, nine, 63. This will be nothing but two by 63. And here you have 0.031 okay. So this is the output that we are specifically getting with respect to uh, the probability of yes with respect to sunny and hot now in order to get the next one and we'll keep this value, I will just show you everything. Now, the next one will be that I need to compute. What is the probability of no given sunny comma hot. Okay. So here I have probability of no multiplied by probability. So the PR p that's one. And the same thing okay. I'm just trying to mention in this way. So what will be the probability of sunny given. No multiplied by what is the probability of hot given. No. See there is a reason why I'm doing this denominator we are not considering. So I'm not going to write. So what is probability of no. It is nothing but five by 14. So here I'm going to write five by 14. See this is quite amazing guys. So clearly so precisely are able to understand what is the probability of sunny given. No. See what is the probability of sunny given. No. What is the probability of sunny given? No three by five. So here I will be having three by five. And what is the probability given hot with respect to no. So here I have two by five. So once I take this and do the computation. So this will be 21705 and five gone. So it will be nothing but three by 35 I hope I hope so I'm doing the calculation right. If there is a small mistake you can let me know. Okay. But at the end of the day you're smart enough. You should be able to understand. So this is the output for probability of no. And this is the output with respect to probability of yes. Right now this output uh is just a probability. But we should try to make this in the form of 100%. Right. So first of all I will go ahead and compute what is the probability of yes given sunny comma hot. Now see see the final conclusion. So here I can basically write what is the probability of this. I got .031 right. So .031. And since I want to make into 100%. So I will divide by 0.031 plus .085. So here you will be able to see that I will be getting 0.27. And this will be nothing but 27 percentage. This step is just done to make sure that we get it. Uh, 100% probability okay. Now the next one. Similarly, if I go ahead and compute the probability of no given sunny comma hot. Now here you have what .085 divided by .085. Sorry, this will not be .085. It will be .031. Right. So this will be .031 plus .085. So if I do the total calculation over here with respect to probability of no I'm going to get somewhere around .73. So this is nothing but seven three 3% okay. Now see this okay. You can obviously say if you get a test data where your outlook feature is sunny and the next feature. Right. If I probably talk about the next feature which was the next feature over here, temperature and your temperature. And your temperature is hot, then what is the output? Your output is nothing, but it is saying 73%. They will not play tennis. They will not play tennis, right? Because no is there and 23%, sorry, 27% they will play tennis, 27% they will play they or he or she will play tennis. Now you know that. What is the highest probability? Obviously 73%. So the outcome for this particular record will be nothing but zero, which is nothing. That person is not going to play the tennis because the probability of no is high. Okay. Not playing tennis. Or not going to play tennis, right? So this is the overall conclusion. And just see every step and why this formula is super important. The derivation of this particular formula is very important because all this X1X2X3 right. It will be converted in the form of features. And this is how naive bias solves any kind of classification problem. Let it be with respect to binary classification or multi-classification. Right. So overall we could see that whenever our new test data was this. Next new test data was this we were able to solve this problem. So. Yes. Uh, I hope you got an idea. Guys, that is the reason I say machine learning theory part. You should really understand implementation. Hardly two lines of code. It is. Okay, so yes, this was it for my side. If you did not understand, please do make sure that you revise it. You will definitely be able to understand. So. Yes. Uh, this was it from my side. I will see you all in the next video. Thank you.

Hello guys. So we are going to continue our discussion with respect to nei biased machine learning algorithm. In our previous video, we have already seen how we can probably use Bayes Theorem and in the name biased machine learning algorithm to solve a binary or multi-class classification problem. In this video, we are going to discuss about the three variants of naive bias that is called as, uh, Bernoulli name bias. Then you have multinomial naive bias and you have Gaussian naive bias. Now all these variants is super important because if you have a different kind of data set right, then which naive bias algorithm you have to probably use from this particular video, you'll be able to derive that specific outcome. Okay. Now let's go ahead and discuss about the first one that is called as Bernoulli name bias. Now what exactly is Bernoulli name bias? It is very much simple, guys. Whenever whenever your features and when I say features, mostly the independent features are following, uh, Bernoulli distribution. I hope. Now you remember what exactly is Bernoulli distribution? We have learned that in statistics. Right then we need to use we need to use Bernoulli nave bias algorithm okay. To solve classification problem. That basically means and you know what is Bernoulli? Bernoulli distribution Bernoulli distribution basically means the outcome that you specifically have in this is either 0 or 1. Right? Tossing a coin. Right. Can it can come out heads or tails. Right. Success or failure. So there will be only two outcomes pass or fail. Right. So if I probably consider with respect to a data set. So in a data set, if I have some examples, let's say I have features like F1, f2, f3 and then I specifically have an output feature. So F1 feature can be like this. Yes. Yes. No. Yes. Right. And similarly F2 feature how it will be. It will be like pass fail pass right fail. So this can be your outcome is either to like see either 0 or 1. Right. So two outcomes are only there. Similarly with respect to F3 it can be a gender feature right. Male female male female right. Like this. And like this it will be continuing. You will be having this repeated feature again and again. And your output can be a binary classification or it can also be a multi classification. So whenever you have this kind of features which is just like kind of a Bernoulli distribution, then you will be able to apply or you should apply Bernoulli naive bias to solve this classification problem. Right. That is the reason it is given a name called as Bernoulli. See, at the end of the day, uh, it's all about simple things. Now, all the statistics that we have used, some or the other way it will be coming to into use right now. See, the last one is called as Gaussian name bias. Okay, now keep on thinking what exactly Gaussian a bias can be if it is Bernoulli a bias over here as first and we have discussed about that, right? Remaining all the techniques to calculate the probability will be almost same. Now coming to the second one, which is called as multinomial naive bias. So let's go ahead and discuss about this one which is second one. Is it nothing but multinomial naive bias? Now this is also a very good algorithm, right. Because for with the help of this, whenever you have your input data in the form of text, then you can probably use multinomial naive bias. Now whenever I say that if your input data is in the form of text and it is a classification problem. So when we say your input data is in the form of text, what does it basically mean? Let's say in your data set we are trying to solve a problem which is called as spam classification. Spam classification. So let's say that you receive a email. So this is my email body okay I need to predict whether this email is spam, whether this email is spam or not spam. So this becomes a classification problem right. So let's say my first message is something like this. Uh, you have $1 million lottery. Million dollar lottery okay. So in this case, obviously the answer will be spam, right. So here let's say that another email is like, uh, I may give something like this. Um, it's like, uh, I'll say my name. Krish, you have done good job. Done good job. So let's say this is my second sentence, and obviously I know that. This mail will not be a spam, so I will write it as ham. Ham basically means not a spam. Now in this scenario, this is my input feature, right? This is my input feature and this is my output feature. Right? So in this my input feature is in the form of text. Right. And when we have this kind of scenario, the first thing that we really need to focus on is that we need to convert this text into a numerical values. The reason we need to convert this into a numerical value is because obviously in this scenario, the model will not be able to understand the sentence. So what we do, how do we convert this into a numerical values. We will be using NLP techniques, natural language processing techniques. Still, we have not started natural language processing guys. But I really want to just give you an idea about it. Later on. As we go ahead with the syllabus, we will be learning about NLP now in natural Language processing in order to convert the sentences into numerical values, we use lot of techniques. The first technique is like bag of words which we say it as bo. Second technique we can go with something called as TF-IDF. So what does this do is that it converts this sentences into some numerical values, which we also say it as vectors okay. But these all techniques is required to do that. Like word two vec like word two vec is also one more technique uh word two vec. And similarly there are a lot of techniques. Even in deep learning there are different different NLP techniques. Now in all these techniques there will be different different formulas, like how we can convert the sentences into numerical values or vectors. So in those formulas, you basically try to find out how many total number of words are there in the sentences? How many total number of unique words are in there in in the sentences, or how many total number of words are there in the entire paragraph? Right. Something like this. So whenever we try to ask this kind of questions, right. And then we try to convert this into a vector by using this formula for this problem statement, whenever your output is spam or ham, we try to use multinomial naive bias, because multinomial naive bias will be very much suitable for this kind of problem statement. Internally, some changes in the formula will be there with respect to understanding this particular implementation. Again, I will not show even show you the practical example right now, but as soon as we start NLP, our first project will be solving based on multinomial naive bias. This all will not work better whenever your input data is in the form of text. Because we are trying to convert our text data into some numerical values. Now coming to the third type okay, so guys, we will learn about Multinomial Naive Bayes. Don't worry about that. Uh, with respect to practical implementation, see in the back end, how does the Naive Bayes algorithm work? It uses Bayes theorem, but for different different data set. This three variants are basically used. Okay, so that optimization wise the accuracy wise it will be able to give you in a better way. Right now coming to the third type. The third type is specifically called as Gaussian naive bias. Now obviously I've told you just think over it. What exactly is Gaussian naive bias? Okay, you know about Bernoulli. So in Gaussian naive bias what do we do in the Gaussian naive Bayes if the. Features are following. Gaussian distribution if the features are following Gaussian distribution. What does Gaussian distribution mean? Bell curve. You remember that in statistics we have learnt it right? Then we use Gaussian name bias for multiclass classification or binary classification. So this variant is only being used when your features are Gaussian distribution. Right. And one of the example that every feature may probably have this kind of distribution. Or it can be a little bit right skewed or left skewed. That will also work well. Right. But even though if it is of a different distribution, then convert this into normal distribution that we have seen how like log normal distribution was there, we have learnt about exponential distribution, right? Exponential distribution, how we can convert a power law into a normal distribution. All those things. Simple transformation formula needs to get applied. Now some of the example. I hope everybody is now familiar with this iris data set. Because in the last problem also we solved this right in our previous algorithms. We use this in iris data set. I had that sepal length petal length petal width sepal width. All these features in short are continuous value. So one important thing with respect to the data set is that here the features are continuous, right? It is continuous. Now when I say continuous I will be having features like age, height, weight okay, let's say here we need to predict whether the whether the customer, whether the person is um, uh, you know, overweight or not. Yes or no. Okay. So I'll be having some values over here like 2538, 22, 24. These are all continuous values. And most of the time this distribution follows this Gaussian distribution only. This will be my mean standard deviation. Mean standard deviation. Right. Something like this. So here I will be having one 7161 5150. And wait I will be having 78, 7560, 35. Right. So yes or no based on the output we will be able to see this. Now whenever you have this continuous features, whenever you have this continuous feature, you specifically use Gaussian name bias. But still you may have a question. Krish, what happens if you have some of the features Bernoulli distribution also over here. Now you need to see that which of the maximum features, like whether the maximum features are having continuous values or whether it is having a Bernoulli distribution. If it is having max Bernoulli distribution, I would suggest always go with this technique called as Bernoulli naive bias. But if you have features that are continuous, and you also have some of the features that have multiple categories inside that, not Bernoulli, then you can go with Gaussian naive bias right at the end of the day, because you need to convert that feature into numerical values also. Right. So this is the basic difference, right. To just understand with respect to this particular variance. But the main important thing that you really need to know is that how the navbar is working, uh, working is done. And for which kind of data set, what variant you really need to apply. And that is what we are going to see in the next video with respect to the practical implementation.