We have been able to break our problem into smaller problem.

We started with the problem of finding which class the data belongs to then we decided lets find probablity of each class.
Then we decided lets use bayes theorem because finding probablity of each class will be higher lets try to find probablity of the input data given the class.
And we have to multiply it with the class probablity as well.
And then using Naive assumption of naive Bayes, what we figured out is that to find out

$$P(X = x/y = a_i) * P(y = a_i) = \prod_{j=1}^{n} P(x^j = x^j/y = a_i) * P(y = a_i)$$

.
Now we are going to find out $P(x^j = x^j)/(y = a_i)$. what should we do to calculate these values.
For this, we assume all our data, all our features are labelled features(Discrete Valued Features).
For eg. lets say we have features, f_1, f_2, f_3.
Lets say f_1 might be taking three different values, a,b,a,c,b,a,c.
Similarly f_2 might be taking three different values, k_1, k_2, k_1, k_2, k_3.
Similarly f_3 might be taking four different values,l_1, l_2, l_1, l_3, l_4.
lets say the f_2 is salary which is marked as low, medium and high.
Lets say f_3 could be repaying loan which can be marked as Yes or No.
So data is some fixed labels. This is for Discrete Data.
So what is the probablity that given x^j class is a_i.
First of all the class is given as a_i, so out of all training data we have, lets focus on a_1.
Whatever we do for a_1 will be applied ofr all other classes.
SO

$$P(x^j = x^j/y = a_i) = \frac{countoftrainingdatasuchthat(X^j = x^j andy = a_1)Thatmeansthejthfeaturevalueisx^j andonlythoserowswhereyisa_1 Iwanttofindhowmanyoftheseh}{countoftrainingdata(y = a_1)}$$

Lets say one of these has desired value that we are looking for, so answer will be 1/3. Because we have 3 possible rowswith class a_1, only one of them has the entry of x^j, so the value is going to be 1/3.
So,

$$P(x^j = x^j/y = a_i) = \frac{Count - in - training - data(X^j = x^j andy = a_1)}{count - in - training - data(y = a_1)}$$

.
Lets take an example

salary	Loan_approved
High	Yes
High	Yes
Mid	Yes
High	Yes
Mid	Yes
Low	Yes
Low	No
Mid	No
Low	No
High	No

Above table is for the Loan Approval.
Lets say what is the Probablity of Salary is High given Loan Approved is No. = $P(salary = High/y = No)$
To do so, we will look at only applications which has y = No.
So we have 4 values in table and fot salary is High, we have only 1 entry of it.
So the Probablity is 1/4.
First we need to count the values of y and then we need to find the x within that rows of y.
Similarly if we want to find the $P(salary = High/y = Yes)$.
First we need to find the the no of y which has Yes and that is 6, now from those 6 rows, we need to check how much salary is High? It is 3 so the answer is 3/6.

Now lets Just talk about given the training Data, how are we going to apply this how are we going to write code to get to the final probablities.
Lets say we have training data, one way to structure this data is because everything is given in $y = a_i$, what we are going to do is we want to create a dictionary, this dictionary on the first level will have $dict[a_i]$, of, which class this basically belongs to? So our top level dictionary will

have all possible different classes, so the dictionary will have the keys $dict[a_1]$, $dict[a_2]$, $dict[a_3]$ .... and so on. And within each of these dictionaries we will have another dictionaries, which will have features.Lets say the features are x^1, x^2, x^3,... and so on. SO the dictionary inside dictionary dict will be $d[x^1]$, $d[x^2]$, $d[x^3]$, ... and so on.
Within these feature dictionary we need to know what are the possible values of these features lets say x^1.
Lets say it has High, Mid, low.
So within x^1, we will have another dictionaries, which will have keys as values of this feature(x^1), (High, mid, low) and these keys will store corrosponding counts of these values of feature.
Also the $dict[a_i]$ will store the count of the class a_i as value also.
And now if we want to find out the probablity $P(X^j = x^j / y = a_i)$, what we simply do is we go to dictionary $dict[a_i]$, wthin that I know I have to look for feature $[X^j]$, within that I know I want to particular label $[x^j]$, which looks like $dict[a_i][X^j][x^j]$, this will give us the count of training data where class is a_i, and the feature X^j value is x^j. We need to divide this by total number of trainng data points where $y = a_i$. What we can do is along with the features, lets store a $[total - count]$ key as well and this will basically store what is the total count of datapoints with particular class a_i. So within a_i, we will have another dict which will have totalcount which now looks like

$$\frac{dict[a_i][X^j][x^j]}{dict[a_i][total - count]}$$

. So whole equation looks like

$$P(X^j = x^j / y = a_i) = \frac{dict[a_i][X^j][x^j]}{dict[a_i][total - count]}$$