Naive Bayes

Shopping Example

Problem statement: To predict whether a person will purchase a product on a specific combination of day, discount, and free delivery using a Naive Bayes classifier.

Under the day, look for variables, like weekday, weekend, and holiday. For any given day, check if there are a discount and free delivery. Based on this information, we can predict if a customer would buy the product or not.

See a small sample data set of 30 rows, with 15 of them, as shown below:

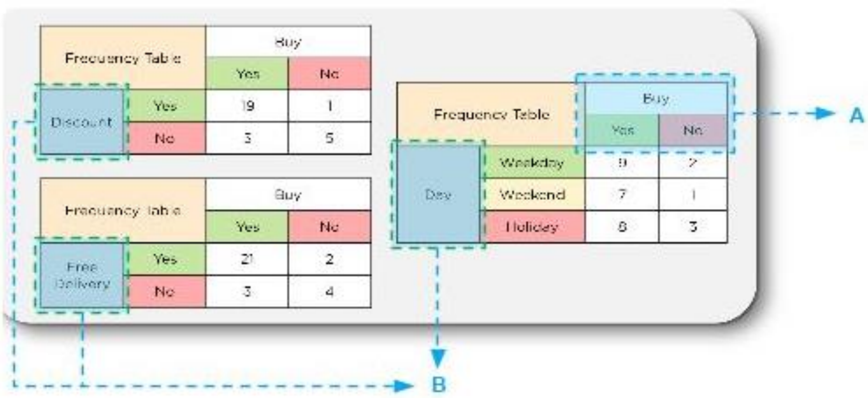
Based on the dataset containing the three input types—day, discount, and free delivery— the frequency table for each attribute is populated.

Frequency Table		Buy	
		Yes	No
Discount	Yes	19	1
	No	5	5

Frequency Table		Buy	
		Yes	No
Free Delivery	Yes	21	2
	No	3	4

Frequency Table		Buy	
		Yes	No
Day	Weekday	9	2
	Weekend	7	1
	Holiday	8	3

For Bayes theorem, let the event 'buy' be A and the independent variables (discount, free delivery, day) be B.



Let us calculate the likelihood for one of the “day” variables, which includes weekday, weekend, and holiday variables.

Frequency Table		Buy		
		Yes	No	
Day	Weekday	9	2	11
	Weekend	7	1	8
	Holiday	8	3	11
		24	6	30

Likelihood Table		Buy		
		Yes	No	
Day	Weekday	9/24	2/6	11/30
	Weekend	7/24	1/6	8/30
	Holiday	8/24	3/6	11/30
		24/30	6/30	

We get a total of:

11 weekdays

Eight weekends

11 holidays

The total number of days adds up to 30 days.

There are nine out of 24 purchases on weekdays

There are seven out of 24 purchases on weekends

There are eight out of 24 purchases on holidays

Based on the above likelihood table, let us calculate some conditional probabilities:

P(B) = P(Weekday)

= 11/30

= 0.37

P(A) = P(No Buy)

= 6/30

= 0.2

P(B | A)

= P(Weekday | No Buy)

= 2/6

= 0.33

P(A | B)

= P(No Buy | Weekday)

= P(Weekday| No Buy) * P(No Buy) / P(Weekday)

= (0.33 * 0.2) / 0.37

= 0.18

The probability of purchasing on the weekday = 11/30 or 0.37

It means out of the 30 people who came into the store throughout the weekend, weekday, and holiday, 11 of those purchases were made on weekdays.

The probability of not making a purchase = 6/30 or 0.2. There's a 20 percent chance that they're not going to make a purchase, no matter what day of the week it is.

Finally, we look at the probability of B (i.e., weekdays) when no purchase occurs.

The probability of the weekday without a purchase = 0.18 or 18 percent. As the probability of (No | Weekday) is less than 0.5, the customer will most likely buy the product on a weekday. Next, let's see how the table and conditional probabilities work in the Naive Bayes Classifier.

We have the frequency tables of all three independent variables, and we can construct the tables for all the three variables.

See the likelihood tables for the three variables below:

Frequency Table		Buy	
		Yes	No
Day	Weekday	3	7
	Weekend	8	2
	Holiday	9	1

Likelihood Table		Buy		
		Yes	No	
Day	Weekday	9/24	2/6	11/30
	Weekend	7/24	1/6	8/30
	Holiday	8/24	3/6	11/30
		24/30	6/30	

Frequency Table		Buy	
		Yes	No
Discount	Yes	19	1
	No	5	5

Frequency Table		Buy		
		Yes	No	
Discount	Yes	19/24	1/6	20/30
	No	5/24	5/6	10/30
		24/30	6/30	

Frequency Table		Buy	
		Yes	No
Free Delivery	Yes	21	2
	No	3	4

Frequency Table		Buy		
		Yes	No	
Free Delivery	Yes	21/24	2/6	23/30
	No	3/24	4/6	7/30
		24/30	6/30	

The likelihood tables can be used to calculate whether a customer will purchase a product on a specific combination of the day when there is a discount and whether there is free delivery. Consider a combination of the following factors where B equals:

- Day = Holiday
- Discount = Yes
- Free Delivery = Yes

Let us find the probability of them not purchasing based on the conditions above.

A = No Purchase

Applying Bayes Theorem, we get P(A | B) as shown:

P (A|B) = P (No Buy | Discount = Yes, Free Delivery = Yes, Day = Holiday)

$$= \frac{P(\text{Discount = Yes} \mid \text{No}) * P(\text{Free Delivery = Yes} \mid \text{No}) * P(\text{Day = Holiday} \mid \text{No}) * P(\text{No Buy})}{P(\text{Discount=Yes}) * P(\text{Free Delivery=Yes}) * P(\text{Day=Holiday})}$$

$$= \frac{(1/6) * (2/6) * (3/6) * (6/30)}{(20/30) * (23/30) * (11/30)}$$

$$= 0.178$$

Similarly, let us find the probability of them purchasing a product under the conditions above.

Here, A = Buy

Applying Bayes Theorem, we get P(A | B) as shown:

$$P(A|B) = P(\text{Yes Buy} \mid \text{Discount} = \text{Yes, Free Delivery} = \text{Yes, Day} = \text{Holiday})$$

$$= \frac{P(\text{Discount}=\text{Yes} \mid \text{Yes}) * P(\text{Free Delivery}=\text{Yes} \mid \text{Yes}) * P(\text{Day}=\text{Holiday} \mid \text{Yes}) * P(\text{Yes Buy})}{P(\text{Discount}=\text{Yes}) * P(\text{Free Delivery}=\text{Yes}) * P(\text{Day}=\text{Holiday})}$$

$$= \frac{(19/24) * (21/24) * (8/24) * (24/30)}{(20/30) * (23/30) * (11/30)}$$

$$= 0.986$$

From the two calculations above, we find that:

Probability of purchase = 0.986

Probability of no purchase = 0.178

Finally, we have a conditional probability of purchase on this day.

Next, normalize these probabilities to get the likelihood of the events:

Sum of probabilities = 0.986 + 0.178 = 1.164

Likelihood of purchase = 0.986 / 1.164 = 84.71 percent

Likelihood of no purchase = 0.178 / 1.164 = 15.29 percent

Result: As 84.71 percent is greater than 15.29 percent, we can conclude that an average customer will buy on holiday with a discount and free delivery.

After understanding how Naive Bayes Classifier works, we can explore its benefits.

Advantages of Naive Bayes Classifier

The following are some of the benefits of the Naive Bayes classifier:

- It is simple and easy to implement
- It doesn’t require as much training data
- It handles both continuous and discrete data
- It is highly scalable with the number of predictors and data points
- It is fast and can be used to make real-time predictions
- It is not sensitive to irrelevant features