

HOME ([HTTPS://WWW.ANALYTICSVIDHYA.COM/](https://www.analyticsvidhya.com/)) BLOG ([HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/](https://www.analyticsvidhya.com/blog/)) ▾



JOBs ([HTTPS://WWW.ANALYTICSVIDHYA.COM/JOBs/](https://www.analyticsvidhya.com/jobs/)) TRAININGS ([HTTPS://WWW.ANALYTICSVIDHYA.COM/TRAININGS/](https://www.analyticsvidhya.com/trainings/))

DISCUSS ([HTTPS://DISCUSS.ANALYTICSVIDHYA.COM/](https://discuss.analyticsvidhya.com/))

LEARNING PATHS ([HTTPS://WWW.ANALYTICSVIDHYA.COM/LEARNING-PATHS-DATA-SCIENCE-BUSINESS-ANALYTICS-BUSINESS-INTELLIGENCE-BIG-DATA/](https://www.analyticsvidhya.com/learning-paths-data-science-business-analytics-business-intelligence-big-data/)) ▾

DATAHACK ([HTTPS://DATAHACK.ANALYTICSVIDHYA.COM/](https://datahack.analyticsvidhya.com/)) ▾ WRITE FOR US ([HTTP://WWW.ANALYTICSVIDHYA.COM/ABOUT-ME/WRITE/](http://www.analyticsvidhya.com/about-me/write/))

DATAHACK SUMMIT 2017 ([HTTPS://WWW.ANALYTICSVIDHYA.COM/DATAHACKSUMMIT/](https://www.analyticsvidhya.com/datahacksummit/)) CONTACT US ([HTTPS://WWW.ANALYTICSVIDHYA.COM/CONTACT/](https://www.analyticsvidhya.com/contact/))

Home (<https://www.analyticsvidhya.com/>) > Business Analytics (<https://www.analyticsvidhya.com/blog/category/business-analytics/>) > 6 Easy Steps to Learn Naive Bayes Algorithm (with code in Python) (<https://www.analyticsvidhya.com/blog/2015/09/naive-bayes-explained/>)

6 Easy Steps to Learn Naive Bayes Algorithm (with code in Python)

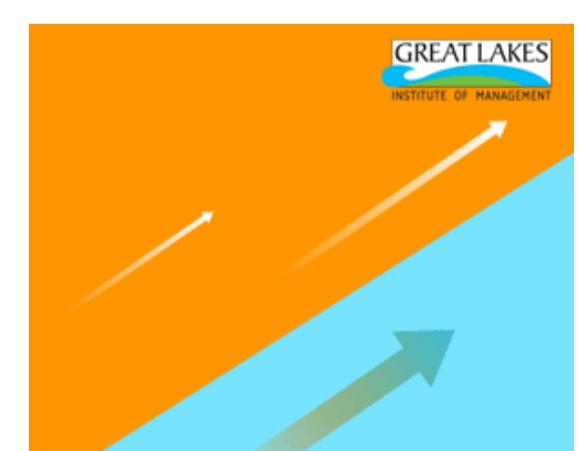
BUSINESS ANALYTICS ([HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/CATEGORY/BUSINESS-ANALYTICS/](https://www.analyticsvidhya.com/blog/category/business-analytics/)) PYTHON ([HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/CATEGORY/PYTHON-2/](https://www.analyticsvidhya.com/blog/category/python-2/))

SHARE ([\) \(\[\\) \\(\\[\\\) \\\(\\\[\\\\)\\\]\\\(https://pinterest.com/pin/create/button/?url=https://www.analyticsvidhya.com/blog/2015/09/naive-bayes-explained/\\\)\\]\\(https://plus.google.com/share?url=https://www.analyticsvidhya.com/blog/2015/09/naive-bayes-explained/\\)\]\(https://twitter.com/home?status=6%20Easy%20Steps%20to%20Learn%20Naive%20Bayes%20Algorithm%20\(with%20code%20in%20Python\)+https://www.analyticsvidhya.com/blog/2015/09/naive-bayes-explained/\)](https://www.facebook.com/sharer.php?u=https://www.analyticsvidhya.com/blog/2015/09/naive-bayes-explained/&t=6%20Easy%20Steps%20to%20Learn%20Naive%20Bayes%20Algorithm%20(with%20code%20in%20Python))



Introduction

Here's a situation you've got into:

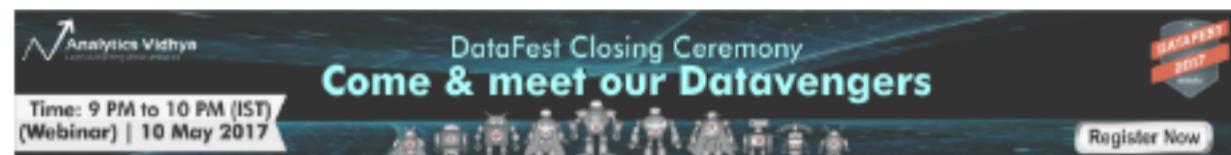


(http://www.greatlearning.in/great-lakes-pgpba?utm_source=avm&utm_medium=avmbanner&utm_campaign=pgpba)

You are working on a classification problem and you have generated your set of hypothesis, created features and discussed the importance of variables. Within an hour, stakeholders want to see the first cut of the model.

What will you do? You have hundreds of thousands of data points and quite a few variables in your training data set. In such situation, if I were at your place, I would have used '**Naive Bayes**', which can be extremely fast relative to other classification algorithms. It works on Bayes theorem of probability to predict the class of unknown data set.

In this article, I'll explain the basics of this algorithm, so that next time when you come across large data sets, you can bring this algorithm to action. In addition, if you are a [newbie in Python](#) (<https://www.analyticsvidhya.com/learning-paths-data-science-business-analytics-business-intelligence-big-data/learning-path-data-science-python/>), you should be overwhelmed by the presence of available codes in this article.



(<https://datahackanalyticsvidhya.com/contest/datafest-closing-ceremony/>)

Table of Contents

1. What is Naive Bayes algorithm?
2. How Naive Bayes Algorithms works?
3. What are the Pros and Cons of using Naive Bayes?
4. Applications of Naive Bayes Algorithm
5. Steps to build a basic Naive Bayes Model in Python
6. Tips to improve the power of Naive Bayes Model

What is Naive Bayes algorithm?

It is a classification technique based on [Bayes' Theorem](#) (https://en.wikipedia.org/wiki/Bayes%27_theorem) with an assumption of independence among predictors. In simple terms, a Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature. For example, a fruit may be considered to be an apple if it is red, round, and about 3 inches in diameter. Even if these features depend on each other or upon the existence of the other features, all of these properties independently contribute to the probability that this fruit is an apple and that is why it is known as 'Naive'.

Naive Bayes model is easy to build and particularly useful for very large data sets. Along with simplicity, Naive Bayes is known to outperform even highly sophisticated classification methods.

Bayes theorem provides a way of calculating posterior probability $P(c|x)$ from $P(c)$, $P(x)$ and $P(x|c)$. Look at the equation below:

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)}$$

Likelihood Class Prior Probability
 Predictor Prior Probability
 Posterior Probability

$$P(c|X) = P(x_1|c) \times P(x_2|c) \times \dots \times P(x_n|c) \times P(c)$$



(https://upgrad.com/data-analytics?utm_source=AV&utm_medium=Display&utm_campaign=DA_AV_Banner&utm_term=DA_AV_Banner&utm_content=DA_AV_Banner)

POPULAR POSTS

- A Complete Tutorial to Learn Data Science with Python from Scratch (<https://www.analyticsvidhya.com/blog/2016/01/complete-tutorial-learn-data-science-python-scratch-2/>)
- Essentials of Machine Learning Algorithms (with Python and R Codes) (<https://www.analyticsvidhya.com/blog/2015/08/common-machine-learning-algorithms/>)
- 7 Types of Regression Techniques you should know! (<https://www.analyticsvidhya.com/blog/2015/08/comprehensive-guide-regression/>)
- 6 Easy Steps to Learn Naive Bayes Algorithm (with code in Python) (<https://www.analyticsvidhya.com/blog/2015/09/naive-bayes-explained/>)
- A Complete Tutorial on Tree Based Modeling from Scratch (in R & Python) (<https://www.analyticsvidhya.com/blog/2016/04/complete-tutorial-tree-based-modeling-scratch->)

(https://www.analyticsvidhya.com/wp-content/uploads/2015/09/Bayes_rule-300x172.png) Above,

- $P(c|x)$ is the posterior probability of class (c , target) given predictor (x , attributes).
- $P(c)$ is the prior probability of class.
- $P(x|c)$ is the likelihood which is the probability of predictor given class.
- $P(x)$ is the prior probability of predictor.

How Naive Bayes algorithm works?

Let's understand it using an example. Below I have a training data set of weather and corresponding target variable 'Play' (suggesting possibilities of playing). Now, we need to classify whether players will play or not based on weather condition. Let's follow the below steps to perform it.

Step 1: Convert the data set into a frequency table

Step 2: Create Likelihood table by finding the probabilities like Overcast probability = 0.29 and probability of playing is 0.64.

Weather	Play
Sunny	No
Overcast	Yes
Rainy	Yes
Sunny	Yes
Sunny	Yes
Overcast	Yes
Rainy	No
Rainy	No
Sunny	Yes
Rainy	Yes
Sunny	No
Overcast	Yes
Overcast	Yes
Rainy	No

Frequency Table		
Weather	No	Yes
Overcast	4	
Rainy	3	2
Sunny	2	3
Grand Total	5	9

Likelihood table			
Weather	No	Yes	
Overcast	4	$\frac{4}{14}$	0.29
Rainy	3	$\frac{3}{14}$	0.36
Sunny	2	$\frac{2}{14}$	0.36
All	5	$\frac{5}{14}$	
		$\frac{9}{14}$	
		0.36	0.64

(https://www.analyticsvidhya.com/wp-content/uploads/2015/08/Bayes_41.png)

Step 3: Now, use Naive Bayesian equation to calculate the posterior probability for each class. The class with the highest posterior probability is the outcome of prediction.

Problem: Players will play if weather is sunny. Is this statement is correct?

We can solve it using above discussed method of posterior probability.

$$P(\text{Yes} | \text{Sunny}) = P(\text{Sunny} | \text{Yes}) * P(\text{Yes}) / P(\text{Sunny})$$

Here we have $P(\text{Sunny} | \text{Yes}) = 3/9 = 0.33$, $P(\text{Sunny}) = 5/14 = 0.36$, $P(\text{Yes}) = 9/14 = 0.64$

Now, $P(\text{Yes} | \text{Sunny}) = 0.33 * 0.64 / 0.36 = 0.60$, which has higher probability.

Naive Bayes uses a similar method to predict the probability of different class based on various attributes. This algorithm is mostly used in text classification and with problems having multiple classes.

in-python/)

A Complete Tutorial on Time Series Modeling in R
(<https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/>)

Understanding Support Vector Machine algorithm from examples (along with code)
(<https://www.analyticsvidhya.com/blog/2015/10/understanding-support-vector-machine-example-code/>)

A comprehensive beginner's guide to create a Time Series Forecast (with Codes in Python)
(<https://www.analyticsvidhya.com/blog/2016/02/time-series-forecasting-codes-python/>)

RECENT POSTS



(<https://www.analyticsvidhya.com/blog/2017/05/beginners-guide-to-data-exploration-using-elasticsearch-and-kibana/>)

Hands on tutorial to perform Data Exploration using Elastic Search and Kibana (using Python)
(<https://www.analyticsvidhya.com/blog/2017/05/beginners-guide-to-data-exploration-using-elastic-search-and-kibana/>)

GUEST BLOG , MAY 15, 2017



(<https://www.analyticsvidhya.com/blog/2017/05/pydata-amsterdam-2017-machine-learning-deep-learning-data-science/>)

What are the Pros and Cons of Naive Bayes?

Pros:

- It is easy and fast to predict class of test data set. It also performs well in multi class prediction
- When assumption of independence holds, a Naive Bayes classifier performs better compare to other models like logistic regression and you need less training data.

- It performs well in case of categorical input variables compared to numerical variable(s). For numerical variable, normal distribution is assumed (bell curve, which is a strong assumption).

Cons:

- If categorical variable has a category (in test data set), which was not observed in training data set, then model will assign a 0 (zero) probability and will be unable to make a prediction. This is often known as "Zero Frequency". To solve this, we can use the smoothing technique. One of the simplest smoothing techniques is called Laplace estimation.
- On the other side naive Bayes is also known as a bad estimator, so the probability outputs from `predict_proba` are not to be taken too seriously.
- Another limitation of Naive Bayes is the assumption of independent predictors. In real life, it is almost impossible that we get a set of predictors which are completely independent.

4 Applications of Naive Bayes Algorithms

- Real time Prediction:** Naive Bayes is an eager learning classifier and it is very fast. Thus, it could be used for making predictions in real time.
- Multi class Prediction:** This algorithm is also well known for multi class prediction feature. Here we can predict the probability of multiple classes of target variable.
- Text classification/ Spam Filtering/ Sentiment Analysis:** Naive Bayes classifiers mostly used in text classification (due to better result in multi class problems and independence rule) have higher success rate as compared to other algorithms. As a result, it is widely used in Spam filtering (identify spam e-mail) and Sentiment Analysis (in social media analysis, to identify positive and negative customer sentiments)
- Recommendation System:** Naive Bayes Classifier and [Collaborative Filtering](https://en.wikipedia.org/wiki/Collaborative_filtering) (https://en.wikipedia.org/wiki/Collaborative_filtering) together builds a Recommendation System that uses machine learning and data mining techniques to filter unseen information and predict whether a user would like a given resource or not

22 must watch talks on Python for Deep Learning, Machine Learning & Data Science (from PyData 2017, Amsterdam)

(<https://www.analyticsvidhya.com/blog/2017/05/pydata-amsterdam-2017-machine-learning-deep-learning-data-science/>)

SUNIL RAY , MAY 11, 2017



(<https://www.analyticsvidhya.com/blog/2017/05/exclusive-interview-with-deepak-vadithala-lead-data-engineer-qlikview-luminary/>)

Inspiring story of Deepak Vadithala – from a Paper delivery boy to a Lead Data Engineer & QlikView Luminary

(<https://www.analyticsvidhya.com/blog/2017/05/exclusive-interview-with-deepak-vadithala-lead-data-engineer-qlikview-luminary/>)

KUNAL JAIN , MAY 9, 2017



(<https://www.analyticsvidhya.com/blog/2017/05/winners-solutions-approach-the-quicksolver-minihack-datafest-2017/>)

Winners solutions & approach: The QuickSolver MiniHack, DataFest 2017

(<https://www.analyticsvidhya.com/blog/2017/05/winners-solutions-approach-the-quicksolver-minihack-datafest-2017/>)

SUNIL RAY , MAY 7, 2017

How to build a basic model using Naive Bayes in Python?

Again, scikit learn (python library) will help here to build a Naive Bayes model in Python. There are three types of Naive Bayes model under scikit learn library:

- Gaussian:** (http://scikit-learn.org/stable/modules/naive_bayes.html) It is used in classification and it assumes that features follow a normal distribution.
- Multinomial** (http://scikit-learn.org/stable/modules/naive_bayes.html): It is used for discrete counts. For example, let's say, we have a text classification problem. Here we can consider bernoulli trials which is one step further and instead of "word occurring in the document", we have "count how often word occurs in the document", you can think of it as "number of times outcome number x_i is observed over the n trials".
- Bernoulli** (http://scikit-learn.org/stable/modules/naive_bayes.html): The binomial model is useful if your feature vectors are binary (i.e. zeros and ones). One application would be text classification with 'bag of words' model where the 1s & 0s are "word occurs in the document" and "word does not occur in the document" respectively.

Based on your data set, you can choose any of above discussed model. Below is the example of Gaussian model.

Python Code

```
#Import Library of Gaussian Naive Bayes model
from sklearn.naive_bayes import GaussianNB
import numpy as np

#assigning predictor and target variables
x= np.array([[-3,7], [1,5], [1,2], [-2,0], [2,3], [-4,0], [-1,1], [1,1], [-2,2], [2,7], [-4,1], [-2,7]])
Y = np.array([3, 3, 3, 3, 4, 3, 3, 4, 3, 4, 4, 4])

#Create a Gaussian Classifier
model = GaussianNB()

# Train the model using the training sets
model.fit(x, y)

#Predict Output
predicted= model.predict([[1,2],[3,4]])
print predicted

Output: ([3,4])
```

Above, we looked at the basic Naive Bayes model, you can improve the power of this basic model by tuning parameters and handle assumption intelligently. Let's look at the methods to improve the performance of Naive Bayes Model. I'd recommend you to go through [this document](http://www.inf.ed.ac.uk/teaching/courses/inf2b/learnnotes/inf2b-learn-note07-2up.pdf) (<http://www.inf.ed.ac.uk/teaching/courses/inf2b/learnnotes/inf2b-learn-note07-2up.pdf>) for more details on Text classification using Naive Bayes.



(http://www.edvancer.in/certified-data-scientist-with-python-course?utm_source=AV&utm_medium=AVads&utm_campaign=AVadsnonfc&utm_content=pythonavad)

GET CONNECTED

9,555
FOLLOWERS

(<http://www.twitter.com/AnalyticsVidhya>)

32,109
FOLLOWERS

(<http://www.facebook.com/AnalyticsVidhya>)

1,826
FOLLOWERS

(<https://plus.google.com/+AnalyticsVidhya>)

Email
SUBSCRIBE

(<http://feedburner.google.com/fb/a/mailverify?uri=analyticsvidhya>)



(<https://www.analyticsvidhya.com/datahacksummit/>)

Tips to improve the power of Naive Bayes Model

Here are some tips for improving power of Naive Bayes Model:

- If continuous features do not have normal distribution, we should use transformation or different methods to convert it in normal distribution.
- If test data set has zero frequency issue, apply smoothing techniques "Laplace Correction" to predict the class of test data set.
- Remove correlated features, as the highly correlated features are voted twice in the model and it can lead to over inflating importance.
- Naive Bayes classifiers has limited options for parameter tuning like alpha=1 for smoothing, fit_prior=[True|False] to learn class prior probabilities or not and some other options (look at detail [here](http://scikit-learn.org/stable/modules/generated/sklearn.naive_bayes.MultinomialNB.html#sklearn.naive_bayes.MultinomialNB) (http://scikit-learn.org/stable/modules/generated/sklearn.naive_bayes.MultinomialNB.html#sklearn.naive_bayes.MultinomialNB)). I would recommend to focus on your pre-processing of data and the feature selection.
- You might think to apply some *classifier combination technique like ensembling, bagging and boosting* but these methods would not help. Actually, "ensembling, boosting, bagging" won't help since their purpose is to reduce variance. Naive Bayes has no variance to minimize.

End Notes

In this article, we looked at one of the supervised machine learning algorithm "Naive Bayes" mainly used for classification. Congrats, if you've thoroughly & understood this article, you've already taken your first step to master this algorithm. From here, all you need is practice.

Further, I would suggest you to focus more on data pre-processing and feature selection prior to applying Naive Bayes algorithm. In future posts, I will discuss about text and document classification using naive bayes in more detail.

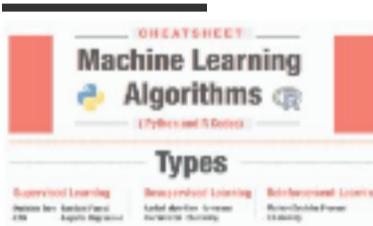
Did you find this article helpful? Please share your opinions / thoughts in the comments section below.

If you like what you just read & want to continue your analytics learning, subscribe to our emails (<http://feedburner.google.com/fb/a/mailverify?uri=analyticsvidhya>), follow us on twitter (<http://twitter.com/analyticsvidhya>) or like our facebook page (<http://facebook.com/analyticsvidhya>).

Share this:

- [!\[\]\(d328bb1c8b293dce97ce8ae48fe06a23_img.jpg\) \(https://www.analyticsvidhya.com/blog/2015/09/naive-bayes-explained/?share=linkedin&nb=1\) 938](https://www.analyticsvidhya.com/blog/2015/09/naive-bayes-explained/?share=linkedin&nb=1)
- [!\[\]\(de0615d88b2098828c20ab3d39ea2ef6_img.jpg\) \(https://www.analyticsvidhya.com/blog/2015/09/naive-bayes-explained/?share=facebook&nb=1\) 713](https://www.analyticsvidhya.com/blog/2015/09/naive-bayes-explained/?share=facebook&nb=1)
- [!\[\]\(6c3f3105811ec6ad9c7c82c1ac88875f_img.jpg\) \(https://www.analyticsvidhya.com/blog/2015/09/naive-bayes-explained/?share=google-plus-1&nb=1\)](https://www.analyticsvidhya.com/blog/2015/09/naive-bayes-explained/?share=google-plus-1&nb=1)
- [!\[\]\(024fb6c2c3e2004cd99a0b34ebd984a9_img.jpg\) \(https://www.analyticsvidhya.com/blog/2015/09/naive-bayes-explained/?share=twitter&nb=1\)](https://www.analyticsvidhya.com/blog/2015/09/naive-bayes-explained/?share=twitter&nb=1)
- [!\[\]\(a018cbc70ca4f8300f04774122af480d_img.jpg\) \(https://www.analyticsvidhya.com/blog/2015/09/naive-bayes-explained/?share=pocket&nb=1\)](https://www.analyticsvidhya.com/blog/2015/09/naive-bayes-explained/?share=pocket&nb=1)
- [!\[\]\(9d74ec3f6ab0c3d629aa0a8c67d82133_img.jpg\) \(https://www.analyticsvidhya.com/blog/2015/09/naive-bayes-explained/?share=reddit&nb=1\)](https://www.analyticsvidhya.com/blog/2015/09/naive-bayes-explained/?share=reddit&nb=1)

RELATED



(<https://www.analyticsvidhya.com/blog/2015/09/full-cheatsheet-machine-learning-algorithms/>)

Cheatsheet - Python & R codes for common Machine Learning Algorithms (<https://www.analyticsvidhya.com/blog/2015/09/full-cheatsheet-machine-learning-algorithms/>)

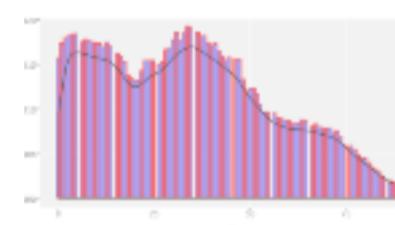
September 14, 2015
In "Business Analytics"



(<https://www.analyticsvidhya.com/blog/2017/01/the-most-comprehensive-data-science-learning-plan-for-2017/>)

The most comprehensive Data Science learning plan for 2017 (<https://www.analyticsvidhya.com/blog/2017/01/the-most-comprehensive-data-science-learning-plan-for-2017/>)

January 16, 2017
In "Business Analytics"



(<https://www.analyticsvidhya.com/blog/2016/09>this-machine-learning-project-on-imbalanced-data-can-add-value-to-your-resume/>)

This Machine Learning Project on Imbalanced Data Can Add Value to Your Resume (<https://www.analyticsvidhya.com/blog/2016/09>this-machine-learning-project-on-imbalanced-data-can-add-value-to-your-resume/>)

September 21, 2016
In "Machine Learning"

TAGS: MACHINE LEARNING (<https://www.analyticsvidhya.com/blog/tag/machine-learning/>), NAIVE BAYES (<https://www.analyticsvidhya.com/blog/tag/naive-bayes/>), PROBABILITY (<https://www.analyticsvidhya.com/blog/tag/probability/>), PYTHON (<https://www.analyticsvidhya.com/blog/tag/python/>)



[Previous Article](#)

[Learn Gradient Boosting Algorithm for better predictions \(with codes in R\) \(<https://www.analyticsvidhya.com/blog/2015/09/complete-guide-boosting-methods/>\)](#)



[Next Article](#)

[Cheatsheet – Python & R codes for common Machine Learning Algorithms \(<https://www.analyticsvidhya.com/blog/2015/09/full-cheatsheet-machine-learning-algorithms/>\)](#)



(<https://www.analyticsvidhya.com/blog/author/sunil-ray/>)

Author

Sunil Ray (<https://www.analyticsvidhya.com/blog/author/sunil-ray/>)

I am a Business Analytics and Intelligence professional with deep experience in the Indian Insurance industry. I have worked for various multi-national Insurance companies in last 7 years.

This article is quiet old now and you might not get a prompt response from the author. We would request you to post this comment on Analytics Vidhya **Discussion portal** (<https://discuss.analyticsvidhya.com/>) to get your queries resolved.

30 COMMENTS



Arun CR says:

REPLY (<https://www.analyticsvidhya.com/blog/2015/09/naive-bayes-explained/?replytocom=94915#respond>)
SEPTEMBER 14, 2015 AT 5:36 AM (<https://www.analyticsvidhya.com/blog/2015/09/naive-bayes-explained/#comment-94915>)

Hi Sunil ,

From the weather and play table which is table [1] we know that frequency of sunny is 5 and play when sunny is 3 no play when suny is 2 so probability(play/sunny) is $3/5 = 0.6$
Why do we need conditional probabiltiy to solve this?

Is there problems that can be solved only using conditional probability. can you suggest such examples.

Thanks,
Arun



matt says:

REPLY (<https://www.analyticsvidhya.com/blog/2015/09/naive-bayes-explained/?replytocom=104496#respond>)



JANUARY 24, 2016 AT 4:42 AM ([HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/09/NAIVE-BAYES-EXPLAINED/#COMMENT-104496](https://www.analyticsvidhya.com/blog/2015/09/naive-bayes-explained/#comment-104496))

Arun,

An example of a problem that requires the use of conditional probability is the Monty Hall problem. https://en.wikipedia.org/wiki/Monty_Hall_problem (https://en.wikipedia.org/wiki/Monty_Hall_problem)

Conditional probability is used to solve this particular problem because the solution depends on Bayes' Theorem. Which was described earlier in the article.



Erdem Karakus says:

MARCH 17, 2016 AT 2:26 PM ([HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/09/NAIVE-BAYES-EXPLAINED/#COMMENT-107604](https://www.analyticsvidhya.com/blog/2015/09/naive-bayes-explained/#comment-107604))

It's a trivial example for illustration. The "Likelihood table" (a confusing misnomer, I think) is in fact a probability table that has the JOINT weather and play outcome probabilities in the center, and the MARGINAL probabilities of one variable (from integrating out the other variable from the joint) on the side and bottom.

Say, weather type = w and play outcome = p.

$P(w,p)$ is the joint probabilities and $P(p)$ and $P(w)$ are the marginals.

Bayes rule described above by Sunil stems from:

$P(w,p) = P(w|p) * P(p) = P(p|w) * P(w)$.

From the center cells we have $P(w,p)$ and from the side/bottom we get $P(p)$ and $P(w)$.

Depending on what you need to calculate, it follows that:

(1): $P(w|p) = P(w,p) / P(p)$ and

(2): $P(p|w) = P(w,p) / P(w)$, which is what you did with $P(\text{sunny},\text{yes}) = 3/14$ and $P(w) = 5/14$, yielding $(3/14) / (14/5)$, with the 14's cancelling out.

The main Bayes take away is that often, one of the two quantities above, $P(w|p)$ or $P(p|w)$ is much harder to get at than the other. So if you're a practitioner you'll come to see this as one of two mathematical miracles regarding this topic, the other being the applicability of Markov Chain Monte Carlo in circumventing some nasty integrals Bayes might throw at you. But I digress.

Best,

Erdem.



John Siano says:

JULY 21, 2016 AT 3:27 AM ([HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/09/NAIVE-BAYES-EXPLAINED/#COMMENT-113751](https://www.analyticsvidhya.com/blog/2015/09/naive-bayes-explained/#comment-113751))

I had the same question. Have you found an answer? Thanks!



Arun CR says:

REPLY ([HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/09/NAIVE-BAYES-EXPLAINED/#COMMENT-94918](https://www.analyticsvidhya.com/blog/2015/09/naive-bayes-explained/#comment-94918))
SEPTEMBER 14, 2015 AT 5:42 AM ([HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/09/NAIVE-BAYES-EXPLAINED/#COMMENT-94918](https://www.analyticsvidhya.com/blog/2015/09/naive-bayes-explained/#comment-94918))

Great article and provides nice information.



Nishi Singh says:

REPLY ([HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/09/NAIVE-BAYES-EXPLAINED/#COMMENT-101806](https://www.analyticsvidhya.com/blog/2015/09/naive-bayes-explained/#comment-101806))
DECEMBER 12, 2015 AT 10:35 AM ([HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/09/NAIVE-BAYES-EXPLAINED/#COMMENT-101806](https://www.analyticsvidhya.com/blog/2015/09/naive-bayes-explained/#comment-101806))

EXPLAINED/#COMMENT-101806)

Amazing content and useful information



Leena says:

REPLY (<https://www.analyticsvidhya.com/blog/2015/09/naive-bayes-explained/?replytocom=105356#respond>)
FEBRUARY 8, 2016 AT 5:59 AM (<https://www.analyticsvidhya.com/blog/2015/09/naive-bayes-explained/#comment-105356>)

I'm new to machine learning and Python. Could you please help to read data from CSV and to separate the same data set to training and test data



Ismael Ezequiel (<https://github.com/IsmaelEzequiel>) says:

REPLY (<https://www.analyticsvidhya.com/blog/2015/09/naive-bayes-explained/?replytocom=117171#respond>)
OCTOBER 15, 2016 AT 12:12 AM (<https://www.analyticsvidhya.com/blog/2015/09/naive-bayes-explained/#comment-117171>)

import pandas as pd

```
person = pd.read_csv('example.csv')
mask = np.random.rand(len(sales)) < 0.8
train = sales[mask]
test = sales[~mask]
```



Sushma horridge says:

REPLY (<https://www.analyticsvidhya.com/blog/2015/09/naive-bayes-explained/?replytocom=106807#respond>)
MARCH 7, 2016 AT 2:44 PM (<https://www.analyticsvidhya.com/blog/2015/09/naive-bayes-explained/#comment-106807>)

very useful article.



RAJKUMAR says:

REPLY (<https://www.analyticsvidhya.com/blog/2015/09/naive-bayes-explained/?replytocom=107493#respond>)
MARCH 16, 2016 AT 1:30 PM (<https://www.analyticsvidhya.com/blog/2015/09/naive-bayes-explained/#comment-107493>)

Very nice....but...if u dont mind...can you please give me that code in JAVA

...



jitesh says:

REPLY (<https://www.analyticsvidhya.com/blog/2015/09/naive-bayes-explained/?replytocom=108955#respond>)
APRIL 5, 2016 AT 3:20 AM (<https://www.analyticsvidhya.com/blog/2015/09/naive-bayes-explained/#comment-108955>)

is it possible to classify new tuple in orange data mining tool??



SPGupta says:

REPLY (<https://www.analyticsvidhya.com/blog/2015/09/naive-bayes-explained/?replytocom=109291#respond>)
APRIL 11, 2016 AT 4:43 PM (<https://www.analyticsvidhya.com/blog/2015/09/naive-bayes-explained/#comment-109291>)

good.



devenir_riche au Maroc (<http://www.coachoutletfactorystore3.com/2015/10/25/jeux-dargent-vs-trading/>) says:

APRIL 14, 2016 AT 5:13 PM (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/09/NAIVE-BAYES-EXPLAINED/#COMMENT-109474)

I am really impressed together with your writing skills as well as with the format to your weblog.

Is that this a paid theme or did you customize it yourself?

Anyway stay up the excellent quality writing, it's uncommon to see a great weblog like this one these days..



รับเช้า-PeteK_Philippe (<https://www.facebook.com/shinbrandname>) says:

APRIL 22, 2016 AT 3:52 AM (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/09/NAIVE-BAYES-EXPLAINED/#COMMENT-109828)

You should be a part of a contest for one of the most useful websites online.

I will highly recommend this blog!



TingTing says: (<https://www.analyticsvidhya.com/blog/2015/09/naive-bayes-explained/?replytocom=109857#respond>)

APRIL 22, 2016 AT 3:18 PM (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/09/NAIVE-BAYES-EXPLAINED/#COMMENT-109857)

Thanks for tips to improve the performance of models, that's really precious experience.



Leanna Partridge says: (<https://www.analyticsvidhya.com/blog/2015/09/naive-bayes-explained/?replytocom=110810#respond>)

MAY 11, 2016 AT 8:18 AM (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/09/NAIVE-BAYES-EXPLAINED/#COMMENT-110810)

Nice piece – Just to add my thoughts , people require a CA OCF-1 , We used a sample document here <http://goo.gl/ibPgs2> (<http://goo.gl/ibPgs2>)



Miguel Batista says: (<https://www.analyticsvidhya.com/blog/2015/09/naive-bayes-explained/?replytocom=110816#respond>)

MAY 11, 2016 AT 9:54 AM (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/09/NAIVE-BAYES-EXPLAINED/#COMMENT-110816)

Hi,

I have a question regarding this statement:

'If continuous features do not have normal distribution, we should use transformation or different methods to convert it in normal distribution.'

Can you provide an example or a link to the techniques?

Thank you,

MB



bd tv (<http://www.bangladeshiliverty.com/>) says: (<https://www.analyticsvidhya.com/blog/2015/09/naive-bayes-explained/?replytocom=111907#respond>)

JUNE 6, 2016 AT 2:45 AM (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/09/NAIVE-BAYES-EXPLAINED/#COMMENT-111907)

Can I simply just say what a comfort to find someone who actually knows what they're talking about over the internet. You certainly know how to bring an issue to light and make it important.

A lot more people ought to read this and understand this side of your story.

I can't believe you aren't more popular because you certainly have the gift.

**nir says:**[REPLY \(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/09/NAIVE-BAYES-EXPLAINED/?REPLYTOCOM=113416#RESPOND\)](https://www.analyticsvidhya.com/blog/2015/09/naive-bayes-explained/?replytocom=113416#respond)[JULY 13, 2016 AT 6:57 PM \(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/09/NAIVE-BAYES-EXPLAINED/#COMMENT-113416\)](https://www.analyticsvidhya.com/blog/2015/09/naive-bayes-explained/#comment-113416)

Great article! Thanks. Are there any similar articles for other classification algorithms specially target towards textual features and mix of textual/numeric features?

**Nick says:**[REPLY \(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/09/NAIVE-BAYES-EXPLAINED/?REPLYTOCOM=114946#RESPOND\)](https://www.analyticsvidhya.com/blog/2015/09/naive-bayes-explained/?replytocom=114946#respond)[AUGUST 19, 2016 AT 6:29 AM \(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/09/NAIVE-BAYES-EXPLAINED/#COMMENT-114946\)](https://www.analyticsvidhya.com/blog/2015/09/naive-bayes-explained/#comment-114946)

great article with basic clarity....nice one

**Catherine says:**[REPLY \(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/09/NAIVE-BAYES-EXPLAINED/?REPLYTOCOM=115295#RESPOND\)](https://www.analyticsvidhya.com/blog/2015/09/naive-bayes-explained/?replytocom=115295#respond)[AUGUST 29, 2016 AT 5:18 PM \(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/09/NAIVE-BAYES-EXPLAINED/#COMMENT-115295\)](https://www.analyticsvidhya.com/blog/2015/09/naive-bayes-explained/#comment-115295)

This article is extremely clear and well laid-out. Thank you!

**pangaxkane_nitin says:**[REPLY \(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/09/NAIVE-BAYES-EXPLAINED/?REPLYTOCOM=115354#RESPOND\)](https://www.analyticsvidhya.com/blog/2015/09/naive-bayes-explained/?replytocom=115354#respond)[AUGUST 31, 2016 AT 8:41 AM \(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/09/NAIVE-BAYES-EXPLAINED/#COMMENT-115354\)](https://www.analyticsvidhya.com/blog/2015/09/naive-bayes-explained/#comment-115354)

ty

**alfiya says:**[REPLY \(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/09/NAIVE-BAYES-EXPLAINED/?REPLYTOCOM=115357#RESPOND\)](https://www.analyticsvidhya.com/blog/2015/09/naive-bayes-explained/?replytocom=115357#respond)[AUGUST 31, 2016 AT 9:26 AM \(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/09/NAIVE-BAYES-EXPLAINED/#COMMENT-115357\)](https://www.analyticsvidhya.com/blog/2015/09/naive-bayes-explained/#comment-115357)

Explanation given in simple word. Well explained! Loved this article.

**Chris Rucker says:**[REPLY \(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/09/NAIVE-BAYES-EXPLAINED/?REPLYTOCOM=115421#RESPOND\)](https://www.analyticsvidhya.com/blog/2015/09/naive-bayes-explained/?replytocom=115421#respond)[SEPTEMBER 1, 2016 AT 7:02 PM \(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/09/NAIVE-BAYES-EXPLAINED/#COMMENT-115421\)](https://www.analyticsvidhya.com/blog/2015/09/naive-bayes-explained/#comment-115421)

The 'y' should be capitalized in your code – great article though.

**Aakash Swamy says:**[REPLY \(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/09/NAIVE-BAYES-EXPLAINED/?REPLYTOCOM=115661#RESPOND\)](https://www.analyticsvidhya.com/blog/2015/09/naive-bayes-explained/?replytocom=115661#respond)[SEPTEMBER 6, 2016 AT 7:24 AM \(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/09/NAIVE-BAYES-EXPLAINED/#COMMENT-115661\)](https://www.analyticsvidhya.com/blog/2015/09/naive-bayes-explained/#comment-115661)

This is the best explanation of NB so far simple and short

**John (<https://floopydugong.github.io/>) says:**[REPLY \(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/09/NAIVE-BAYES-EXPLAINED/?REPLYTOCOM=115863#RESPOND\)](https://www.analyticsvidhya.com/blog/2015/09/naive-bayes-explained/?replytocom=115863#respond)[SEPTEMBER 9, 2016 AT 5:20 PM \(HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/09/NAIVE-BAYES-EXPLAINED/#COMMENT-115863\)](https://www.analyticsvidhya.com/blog/2015/09/naive-bayes-explained/#comment-115863)

Great article! Really enjoyed it. Just wanted to point out a small error in the Python code.

Should be a capital "Y" in the predict like so : model.fit(x, Y)

Thanks!

**Adnan says:**REPLY (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/09/NAIVE-BAYES-EXPLAINED/?REPLYTOCOM=116232#RESPOND)SEPTEMBER 20, 2016 AT 5:12 AM (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/09/NAIVE-BAYES-EXPLAINED/#COMMENT-116232)

Is this dataset related to weather? I am confused as a newbie. Can you please guide?

**bh says:**REPLY (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/09/NAIVE-BAYES-EXPLAINED/?REPLYTOCOM=117314#RESPOND)OCTOBER 19, 2016 AT 5:40 PM (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/09/NAIVE-BAYES-EXPLAINED/#COMMENT-117314)

best artical that help me to understand this concept

**Richard says:**REPLY (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/09/NAIVE-BAYES-EXPLAINED/?REPLYTOCOM=117495#RESPOND)OCTOBER 25, 2016 AT 7:52 PM (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/09/NAIVE-BAYES-EXPLAINED/#COMMENT-117495)

Am new to machine learning and this article was handy to me in understanding naive bayes especially the data on weather and play in the table.

Thanks for sharing keep up

**Lisa says:**REPLY (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/09/NAIVE-BAYES-EXPLAINED/?REPLYTOCOM=118208#RESPOND)NOVEMBER 10, 2016 AT 8:29 AM (HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/09/NAIVE-BAYES-EXPLAINED/#COMMENT-118208)

Thanks to you I can totally understand NB classifier.

LEAVE A REPLY

Your email address will not be published.

Comment
Name (required)
Email (required)
Website

SUBMIT COMMENT

ANALYTICS VIDHYA

About Us (https://www.analyticsvidhya.com/about-me/)	Blog (http://analyticsvidhya.com/blog)
Team (https://www.analyticsvidhya.com/about-team)	Discuss (https://discuss.analyticsvidhya.com)
Volunteers (https://www.analyticsvidhya.com/av-volunteers)	Learning Paths (http://analyticsvidhya.com/learning-paths)
Careers (https://www.analyticsvidhya.com/about-career-analytics-vidhya)	business-analytics business-intelligence-big-data/ DataHack (https://datahack.analyticsvidhya.com)
Contact (https://www.analyticsvidhya.com/contact)	Events (https://analyticsvidhya.com/blog/category/events/)
Jobs (https://www.analyticsvidhya.com/jobs)	Jobs (https://www.analyticsvidhya.com/jobs/)

DATA SCIENTISTS**COMPANIES****SOCIAL MEDIA**

Advertise with Us (https://www.analyticsvidhya.com/contact)	 9,555 FOLLOWERS
Recruit using AV (https://www.analyticsvidhya.com/jobs/company)	 32,189 FOLLOWERS
/signup/ Hackathon (https://datahack.analyticsvidhya.com/)	 1,826 FOLLOWERS
DataFest-2017 (https://www.analyticsvidhya.com/datafest-2017)	 Email SUBSCRIBE