IJARSCT

ISSN (Online) 2581-9429

IJARSCT
Impact Factor: 5.731

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

Volume 10, Issue 1, October 2021

Study on Naive Bayes Algorithms and How They Work in Real World

Saurabh Subhashchand Yadav
Post Graduate Student M.Sc., Department of Information Technology
Sir Sitaram and Lady Shantabai Patkar College of Arts and Science, Mumbai, India
saurabhsyadav1999@gmail.com

Abstract*:* *A classification algorithm based on the Bayes theorem with strong and naive independence assumptions is known as Naive Bayes. It makes learning easier by understanding that features are unaffected by the class. This paper discusses the concept of the nave Bayes algorithm, hidden nave Bayes, text classification, classic nave Bayes, and machine learning. Finally, some applications of nave Bayes, as well as its benefits and drawbacks, are explored for a better understanding of the algorithms.*

Keywords: Big Data, Naive Bayes, Data Mining, Recommend System

I. Introduction

Naïve Bayesian decision theory has a subset called Bayes. Because the formulation contains some naive assumptions, it is called naive. Python's text-processing skills are applied to break a document into a vector. Text can be classified using this method. It is possible to convert the classifications into a human-readable format. Along with conditional independence, overfitting, and Bayesian approaches, it is a common classification method.

Easy implementation, Naive Bayes has a surprising ability to classify documents. An initial reason for the conditional independence assumption is that if the document is about politics, this is solid proof of the types of other words found in the document. In this sense, Naive Bayes is a good classifier with low storage and quick training.

The supervised classifying problem is a worldwide topic, and most of the approaches for building such rules have been developed. It is relatively simple to set up, and no complicated, repeating parameter estimating strategies are required.

Finally, it frequently performs impressively: it may not be the best possible classifier in any given application, but it can typically be counted on to be stable and perform well.

It is relatively simple to set up, and no complicated, repeating parameter estimating strategies are required. This implies that it should be used on large data sets.

II. Naive Bayes Classifier

It is relatively simple to set up, and no complicated, repeating parameter estimating strategies are required. This implies that it should be used on large data sets. The Nave Bayes classifier is a probabilistic classifier based on Bayes' theorem, which maintains that each feature helps to the target class independently and equally.

The NB classifier believes that each feature is independent of the others and that each feature contributes equally to the probability of a sample belonging to a certain class. The NB classifier is easy to use and fast to compute, and it works well with huge datasets with a lot of factors.

It's mostly applied in text classification with a large training dataset. It's a probabilistic classifier, which means it estimates based on an object's probability. Spam filtration, sentiment analysis, and article classification are all popular uses of the Nave Bayes Algorithm.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

P(A|B) is Posterior probability: Probability of hypothesis A on the observed event B.
P(B|A) is Likelihood probability: Probability of the evidence given that the probability of a hypothesis is true.
P(A) is Prior Probability: According to observing the data, the probability of the hypothesis.
P(B) is Marginal Probability: Probability of evidence .

III. Naive Bayes Applications

3.1 Text Classification

For text classification, it is applied as a probabilistic learning method. When it comes to text document classification, the Naive Bayes classifier is one of the most well-known algorithms. It determines if a text document belongs to one or more categories (classes).

3.2 Spam Filteration

It's a text classification example. This has become an useful way to tell the difference between spam and authentic email. Bayesian spam filtering is used by a number of modern email providers. This approach is used by many server-side email filters, including DSPAM, Spam Bayes etc.

3.3 Sentiment Analysis

It can be used to detect whether tweets, comments, and reviews are bad, good, or balanced in content.

3.4 Recommendation Systems

The Naive Bayes algorithm in combination with collaborative filtering is used to build hybrid recommendation systems

which help in predicting if a user would like a given resource or not.

3.5 Steps to implement Naïve Bayes

- Step 1: Pre process the Data .
- Step 2: Fit Naïve Bayes into the training set .
- Step 3: Predicting the result .
- Step 4: Check test accuracy of the result .
- Step 5: Visualize the test set result .

IV. Advantages and Disadvantages

4.1 Advantages

- It's a straightforward algorithm to understand and implement.
- This algorithm is faster than many other classification algorithms at identifying classes.
- It's simple to train with a small dataset.

4.2 Disadvantages

- In Naive Bayes, all variables (or attributes) are assumed to be independent, which is rare in real life. This limits the algorithm's use in real-world scenarios.
- You shouldn't take its probability outputs seriously because its estimations can be off in some situations.

V. Types of Naïve Bayes Model

5.1Gaussian Model

The Gaussian model assumes that variables are normally distributed. If variables take continuous values rather than discrete values, the model assumes that these values are chosen from a Gaussian distribution.

5.2 Multinomial Model

When the data is multinomial distributed, the Multinomial Nave Bayes classifier is used. It is primarily used to solve document classification issues, which involves determining which category a document belongs to, such as Sports, Politics, or Education. The predictions in the classifier are based on the frequency of terms.

5.3 Bernoulli Model

The Bernoulli classifier works in a similar way to the Multinomial classifier, except that the parameters are independent Booleans values. For example, determining whether or not a specific word appears in a document. This model is also well-known for tasks involving document classification.

VI. Conclusion

This paper gives a survey of the Naive Bayes Algorithm, discussing improved Naive Bayes text classification, Spam filtration, Sentiment analysis, and Recommendation System as some of the algorithm's key applications. It also has certain issues, such as how to solve the Zero Based On the weighted Problem. A class conditional probability estimation is a critical problem in the naive Bayes model, and core interpolation is a popular method for using it.

References

[1]. I. Rish, "An Empirical Study of the Naïve Bayes Classifier," no. January 2014.

[2]. J. Ren, S. D. Lee, X. Chen, B. Kao, R. Cheng, and D. Cheung, —Naive Bayes Classification of Uncertain Data,‖ no. 60703110.

[3]. Tom M. Mitchell. Machine Learning.McGraw-Hill, 1997 .

[4]. Harry Zhang, —The Optimality of Naive Bayes‖, American Association for Artificial Intelligence, 2004.

[5]. Toon Calders, SiccoVerwer, —Three naive Bayes approaches for discrimination-free classification‖, Data Min Knowl Disk, 2010.

[6]. Siddharth Banga, SakshamMongia, Vaibhav Tiwari, Mrs. SunitaDhotre,-Regression and Augmentation Analytics on Earth's Surface Temperature‖, IJCST, 2017.

[7]. https://www.upgrad.com/blog/naive-bayes-classifier/

[8]. https://www.javatpoint.com/machine-learning-naive- bayes-classifier

[9]. Liangxiao Jiang, Harry Zhang, and ZhihuaCai, —A Novel Bayes Model, Hidden Naive Bayes‖, IEEE Transaction on Knowledge and Data Engineering, Vol 21, No. 10, pp. 1361 – 1371 .

[10]. W. Zhang and F. Gao, —Procedia Engineering An Improvement to Naive Bayes for Text Classification,‖ vol. 15, pp. 2160–2164, 2011.

DOI: 10.48175/568