iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 7 **Issue:** IV **Month of publication:** April 2019

DOI: https://doi.org/10.22214/ijraset.2019.4032

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 7 Issue IV, Apr 2019- Available at www.ijraset.com

A Literature Review on Hierarchical Naive Bayes Classifier

Ronak Chavan[1], Manish Rana[2]
[1]*Department of Computer Engineering, Thakur College of Engineering and Technology, Mumbai*
[2]*Assistant Professor, Department of Computer Engineering, Thakur College of Engineering and Technology, Mumbai*

Abstract: ***In machine learning, classification of data has an area of many problems Naïve Bayes Model is one of the simplest and widely used models for classification. However, one of the most immanent issue with this classifier is that the attributes used to describe an instance are assumed to be independent of the given class. When this assumption is flouted then the accuracy is decreased due to interaction omission. In this review paper, we check a new model known as Hierarchical Naïve Bayes model and how it gives better and accurate results than the Naïve Bayes model.***
Keywords: Classification, Naïve Bayes models, Hierarchical model, Classifier, Hierarchical Naïve Bayes classifier

I. INTRODUCTION

Classification is the predictive technique of the category to which the data belongs. It is a technique of predicting the class of an instance from a set of attributes describing that instance. Naïve Bayes model is a relatively easy model of classification, this model assumes that all the attributes are independent given the value of class variables. This is not a single algorithm but a group of algorithms which consists of three classifiers. In a classification problem, a set of labelled training set is given to the algorithm, these are defined by some features. Based on the features, the algorithms aim is to evaluate the label of unknown labelled objects. In a flat classification problem, c an element of C is will get assigned to every test example e. This will output only a single class for every test example. One more issue with the Naïve Bayes classifier is that it assumes that all attributes are independent, in real world problem this assumption gets violated. Thus there arrived a need to extend the Naïve Bayes algorithm in order to handle the dependencies between the attributes.

II. NAÏVE BAYES CLASSIFIER

The Naïve Bayes classifiers is family of classifiers which are based on the Bayes Theorem. It is among the easiest and consistently performing classifiers. Instead of calculating the values of every attributes, they're assumed to be independent [1]

The Bayes Theorem is mathematically stated as $P(A|B) = \frac{P(B|A)*P(A)}{P(B)}$

1) P(A) is the prior probability of A.
2) P(A|B) is the posteriori probability.
3) P(B) is the probability of the data.

Now we calculate the independence among the features. If A and B are independent events then P(A,B) = P(A) * P(B)
Thus we get the equation as:
A= $\max_A P(A)\prod_{i=1}^{n} P(B_i|A)$
a) P(A) is the class probability of A.
b) $P(B_i|A)$ is the conditional probability.

III.ISSUES OF NAÏVE BAYES CLASSIFIER

The Naïve Bayes classifier assumes that all attributes are independent, this is called as Conditional Independence. Due to this assumptions, the information that is overlapping is counted twice by the classifier. This assumption is inefficient and can give inaccurate results as real world data doesn't adhere to this assumption. Another issue emerges for continuous features. It is common to utilize a binning method to make them discrete, yet in the event that you are not cautious you can discard a great deal of data. Plausibility is to utilize Gaussian appropriations for the probabilities.

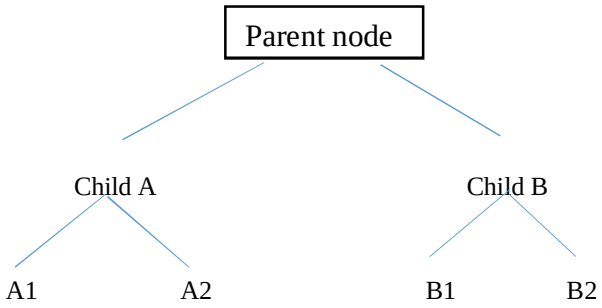
Another issue occurs because of information shortage. You have to measure probability esteem by a frequented method for any conceivable estimate. This can lead to the probabilities of 0 or 1 resulting in numerical insecurity and more regrettable results [1]. Your probabilities have to be smoothened in this situation or force some on your information earlier, anyway, you might argue that the subsequent classifier is no longer naive.

©IJRASET: All Rights are Reserved

IV. HIERARCHICAL CLASSIFICATION

Hierarchical classification is distinguished by two type i.e. local model and global model. In local model, a local classifier is created for every parent node or for each class node a local binary classifier. In the erstwhile case, the goal of the classifier is to differentiate between the children of the equivalent node. In the last case, each binary classifier predicts whether an example belongs to its corresponding class or not. A local view of the problem is created by these approaches in both of these cases. In testing phase, top-down prediction strategy is used with these approaches, though there is a difference in creating and training the classifying algorithm [4]. The decision on which class is anticipated at the current level is based on the class predicted at the prior (parent) level for each level of the hierarchy [1]. The main downside of the local approach with the top - down class prediction approach is that a classification error at a high hierarchy level spreads through all of the wrongly assigned class descendant nodes.

A single (relatively complex) classification model is constructed from the training set in the global model approach, taking into consideration the class hierarchy as a whole while running a single algorithm. Each test example is classified by the induced model when used during the testing phase, a process that can allocate classes to the test example at possibly each level. In addition, in most approaches a top - down class prediction approach is used in the local model. Single global model will be much easier than local model hierarchy, even though global models are more complex than the local models [4].



Hierarchical class

V. HIERARCHICAL NAÏVE BAYES CLASSIFIER

Hierarchical Naïve Bayes model is a special class of Bayesian networks. The variables are partitioned into disjoint sets in this tree-shaped Bayesian network, in this the variables are discrete. This model extends the Naïve Bayes model to deal with the hierarchical classification problem that plagues the standard Naïve Bayes model. The classification model is said to be global since it is built by taking into account all classes in the hierarchy – rather than building a number of local classification models. We also supplement the global-model NB by using a notion of "practicality" by taking into account the gravity of the prediction. Models [1]. This model introduces latent variables to ease some of the independent statements. Results show that this specific model can drastically improve classification accuracy. Project introduces efficiency of HNB Classifier used on a particular dataset dividing it into different polarity segments [1].

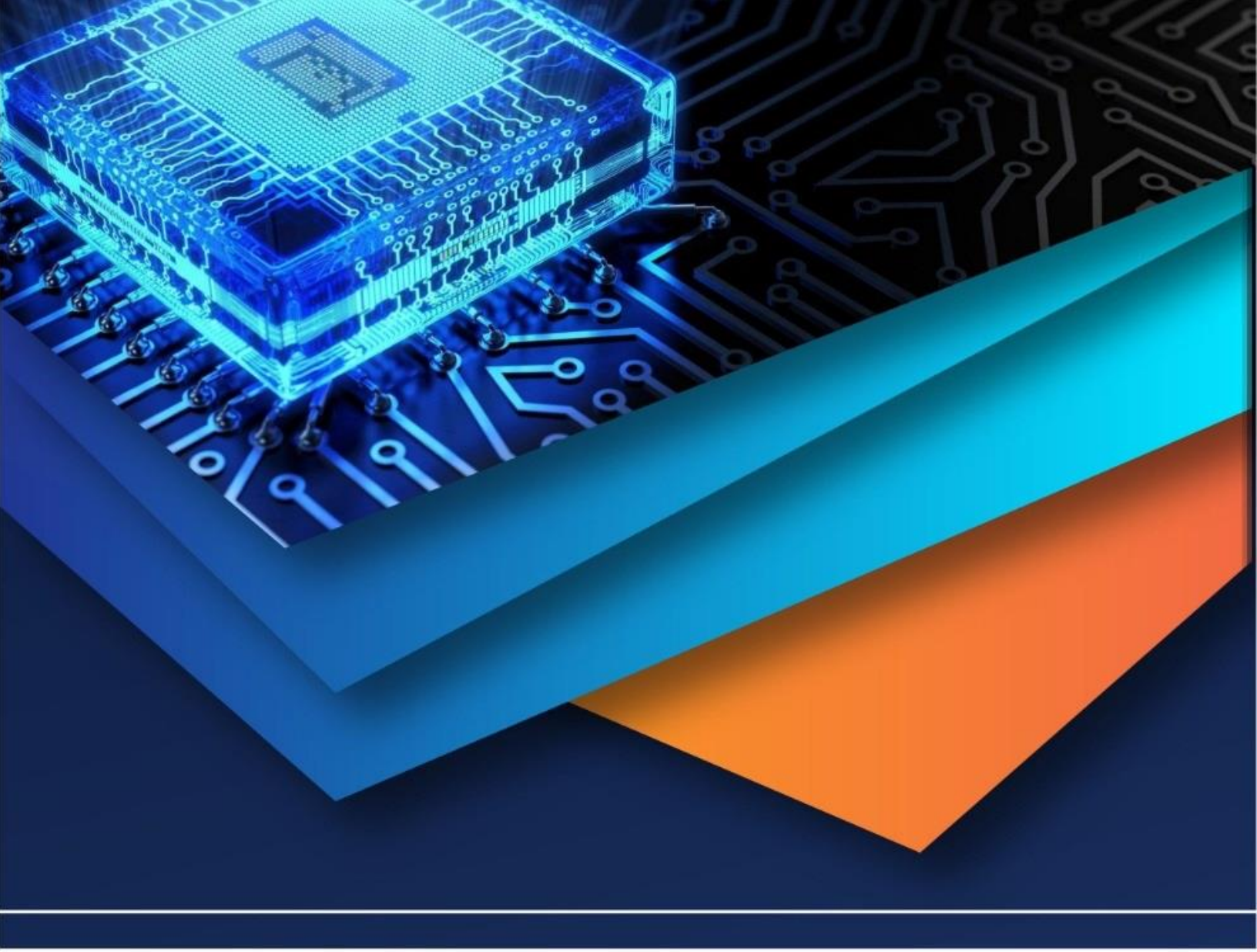
VI. CONCLUSIONS

We have used HNB models for classification, and through experiments it is evident that the HNB classifiers offer significantly better results than those of other generally used classification methods. Moreover, a number of existing tools may be able to further improve the classification accuracy. Finally, the proposed learning algorithm also provides a clear semantics for the latent structure of a model. This allows a decision maker to easily deduce the rules which govern the classification of some instance hence; the semantics may also increase the user's confidence in the model.

REFERENCES

[1] Helge Langseth Thomas D. Nielsen "Classification using Hierarchical Naïve Bayes models" May 2006.
[2] Francesca Demichelis, Paolo Magni, Paolo Piergiorgi, Ricardo Bellazzi "A hierarchical Naïve Bayes Model for handling sample heterogeneity in classification problems: an application to tissue microarrays", November 2006.
[3] https://www.sciencedirect.com/science/article/pii/S0167923611000327
[4] Carlos N. Silla Jr. and Alex A. Freitas "A Global-Model Naive Bayes Approach to the Hierarchical Prediction of Protein Functions" 2009

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)