Journal of Theoretical and Applied Information Technology
15th February 2026. Vol.104. No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

ANALYSIS ACCURACY OF CLASSIFICATION FOR MULTI-EVIDENCE DATASET CONTAINING DISCREATE AND CONTINUOUS VALUE USING CATEGORICALNB AND GAUSSIANNB

DR. VIJAY KUMAR VERMA[1]
[1]Associate Professor Shri Vaishnav Vidyapeeth Vishwavidyalaya Indore M.P. India
[1]drvijaykumarverma20@gmail.com
ORCID 0000-0001-5442-0113

ABSTRACT

Machine Learning is one of the emerging fields in computer science, and Bayes' theorem can be applied to predict class labels precisely and accurately. Naïve Bayes classification, specifically the Categorical Naïve Bayes (CategoricalNB or CNB) classifier, is a simple probabilistic model based on Bayes' theorem and is highly effective in reducing computational cost, as conditional probabilities can be easily estimated from data. In this paper, we employ both Categorical Naïve Bayes (CNB) and Gaussian Naïve Bayes (GNB) classifiers. The CNB classifier is widely used for binary and multi-class classification problems in machine learning and performs well when the attribute values are categorical or discrete. However, it has certain limitations when handling continuous data. In real-world applications, datasets often contain attributes in mixed formats, including both discrete and continuous values. For continuous attributes, CNB requires discretization, whereas GNB can be directly applied to continuous data by assuming a Gaussian distribution. In this study, we use CNB and GNB to classify a dataset containing both discrete and continuous attributes and compare their classification accuracy. A real-life dataset obtained from the HDFC Bank home loan department was used for experimentation. The dataset consists of 6,000 records with 19 attributes. Experimental results show that the performance of the Gaussian Naïve Bayes classifier is superior to that of the Categorical Naïve Bayes classifier. However, a limitation of GNB is its assumption that continuous attributes follow a normal distribution.

Keywords. *Classification, Machine learning, Naive Bays, Continues, Discreate, Gaussian Naïve, Accuracy*

1. INTRODUCTION

Bayes' theorem, also known as Bayes' rule or Bayes' law, helps to calculate the probability of an event based on random data. It allows us to compute the probability of an event occurring given that another event has already occurred. We can estimate the precision of values and compute conditional probability (CP). Although it is a probabilistic calculation, it is relatively easy to compute the conditional probability of events where human perception often fails.

Bayes' theorem is useful in several real-life applications such as the financial industry, healthcare and medical diagnosis, research, surveys, and the aeronautical sector, among others. It provides a method in which the value of $P(X \mid Y)$is known from a given training dataset, and we can calculate $P(Y \mid X)$. This is achieved by simply interchanging the positions of X and Y, where X represents the feature vector and Y represents the response or class label.

When variable Y has more than two categories, we need to calculate the probability for each class of the variable Y [1,2,3].

Bayes theorem or Bayes rule or Bayes Law helps to calculate the probability of an event with random data. We can compute the probability of an event happening although others have already happened. We can use estimate precision of values and compute the conditional probability (CD). Although it is hypocritical calculation, it is easy to calculate the conditional probability of events where perception often fails. It is useful for several real-life areas like financial industries, health and medical, research, survey, aeronautical sector, etc. It provides a way where value for P(X|Y) is known from given training dataset and calculate P(Y|X). Wes can be

simply replacing position of X and Y, with the X feature and response. When variable Y has more than 2 categories, we need to calculate the probability of each class of variable Y [3,4,10].

Bays Rule is way to go $P\left(\frac{X}{Y}\right)$ to find $P\left(\frac{X}{Y}\right)$

$$P\left(\frac{X}{Y}\right) = \frac{P(X \cap Y)}{P(Y)}$$

Bays Rule is way to go $P\left(\frac{X}{Y}\right)$ to find $P\left(\frac{X}{Y}\right)$

$$P\left(\frac{X}{Y}\right) = \frac{P(X \cap Y)}{P(Y)}$$

Known P(Evidence |Outcome)
(Known from training data)

$$P\left(\frac{Y}{X}\right) = \frac{P(X \cap Y)}{P(X)}$$

With the help of this theorem, we can compute the probability value of Y when the value of X is given. However, in real-life datasets, we often have multiple independent attributes. When the number of features increases, it becomes necessary to extend Bayes' rule. In the presence of multiple independent feature variables, the Bayes' rule can be **simplified**, leading to the Naïve Bayes formulation **[11, 12, 16]**.

$$P(Y = k|X) = \frac{P(X|Y = k) * P(Y = k)}{P(X)}$$

Where k is a class of Y
Becomes Naïve Bayes

$$P(Y = k|X_1 \dots \dots X_n)$$
$$= \frac{P(X_1|Y = k) * P(X_2|Y = k) \dots . * P(X_n|Y = k) * P(}{P(X_1) * P(X_2) \dots \dots \dots * P(X_n)}$$

$$probabilty\ of\ Outcome|Evidence$$

$$= \frac{Probability\ of\ Likeihood\ of\ evidence * Pric}{Probability\ of\ Evidence}$$

In the above equation the left-hand side is the posterior probability. In the numerator there are two terms in RHS. The first is 'Likelihood Evidence'.

2. MATERIALS AND METHODS

Features selection provides directions to minimize the input features to model and only used relevant features and receiving noise free data. This process comprises or rejects significant features without changing them. It supports noise removing of data and reducing the size of data input [12,13,15,].

There are three categories of Naïve Bays is available

A. *Multinomial NB*
B. *Bernoulli NB*
C. *Gaussian NB*

2.1 Multinomial NB

Multinomial NB is based on Bayes' theorem and specially used for documents categorizing for multiple classes, Natural Language Processing (NLP) is most common example. It is also useful for spam detection, analysis of sentiment, and classification of text. We can easily compute the probability of class by using *Multinomial NB* for given features.

Mathematical statement for Bayes' theorem:

$$p(y/x) = (p(x/y) * p(y))/(x)$$

where:

- $P(y|x)$ find probability of y when given x.
- $P(x|y)$ find probability of x when given y.
- $P(y)$ prior probability of y.
- $P(x)$ probability of x.

Text classification includes input documents, and classes with different labels for which we are going to classify. The objective is to find the class which has the highest probability for giving the input. Benefits and limitations of Multinomial Naive Bayes

Benefits:

1. Suitable for sparse data with high-dimensionality, useful for text classification where large number of classes are given.

Disadvantages:

1. Assuming features are independence, in real life it is not possible true.
2. Large amount of training data is required for higher accuracy.
3. Suffering from overfitting.

2.2 Bernoulli Naive Bayes

Bernoulli NB is a probabilistic model used to predict belonging for a particular class. It is used for classification of binary classes[15,16,].

It assumes that each feature is independent from one another and follows a Bernoulli distribution. Bernoulli NB first calculates the prior probability, likelihood for the feature and at last posterior probability for each class. Mathematical equations for Bernoulli NB is given as follows [17,18,19].

$$P(y/x) = p(y) * \pi P(x_i/y)$$

Where:

- $P(y|x)$ Find posterior probability for class y
  given features x
- $P(y)$ Prior probability for class y
- $P(x_i/y)$ likelihood for feature i given class

Bernoulli NB is a easy, and effective classifier and appropriate for classification of text. Benefit and limitation Bernoulli Naive Bayes are

Benefits:

1. Computationally efficient for large datasets.
2. It is especially useful for text classification problems.

Disadvantages:

1. Unable to handle continuous data.

3. PROBLEMS WITH THE CATEGORICAL NAÏVE BAYES (CNB) CLASSIFIER

Some of the major problems associated with the Categorical Naïve Bayes (CNB) classifier are as follows:

1. **Predefined classes:** Since the classes are predefined, the classifier must determine and explain the characteristics of each class accurately.
2. **Feature independence assumption:** The classifier assumes that all features are independent of each other and have equal importance in the classification process.
3. **Unrealistic independence in real-world data:** The Naïve Bayes classifier assumes class-conditional independence, which is rarely true in real-life datasets where features are often correlated.
4. **Handling continuous attributes:** Attributes with continuous values must be converted into discrete values. The selection of discretization ranges is user-dependent and may vary, which can significantly affect classification accuracy and other performance measures.

There are several other important issues that need to be addressed while using Naïve Bayes classifiers, which are discussed in the literature [20, 21].

There are other important issues which are needed to handle during naïve bays classifiers these are given below [19,20,21]

1. **Zero frequency problem**
2. **Missing values**
3. **Numeric attributes**

- **Zero frequency problem: -** When an attribute value doesn't occur no matter how probable the other values are in this can need to use Laplace correction.
- **Missing values: -** Missing values have no effect on result.
- **Two classes have different distribution: -** When one class has normal distribution, but other classes have no normal distribution.

4. RELATED WORKS

Unggul Widodo et al [3] proposed Experimental Study for analysis of Sentiment using Gaussian NB". They used customer reviews data from Yelp (foods)and Amazon (products). They want to analyzed customer reviews. So fast reading of customer reviews they used using sentiment analysis. Quality data is used to improve the accuracy of the results. **Nafizatus Salmi** et al [4] proposed Models for Predicting the Colon Cancer using Naïve Bayes Classifier. They showed that the model has good accuracy, precision, recall, f1 − *score* for classifying patients data suffering from colon cancer. The model predicts and is able to make higher accuracy and less complexity. **Allsela Meiriza** et al [5] Predicted student graduation with Naïve Bayes Classifier. With the prediction of the students, graduation Information System makes policies for future improvement. They used WEKA software for data processing. They showed that from the Information Systems Study Computer Science Faculty the prediction accuracy was 97, 6378. **Shikha Agarwal** et al [6] proposed Hybrid model of both Naive Bayes and Gaussian Naive Bayes for Classification. They used Hadoop distributed computing and map reduce programming method for Hybrid model for classification. **Amitha, Merin** et al [7] proposed a Recommendation system for Medication with Gaussian Naïve Bayes method. This system provides suggestions according to health status. recommendation system is responsible for recommending medicines based on the symptoms. **Marzuki Ismail** et al [8] proposed a Comparative analysis for Health-Related Classification Task using Naive Bayesian Techniques. They used integrated approach to opinion mining and sentiment analysis for diseases prediction. Three algorithms were used to show which algorithm is better to classify the given datasets. **Jiajie Shen** et al [9] proposed recognition system for human activity. They applied Gaussian NB with smart environment sensor data. By the experiment they showed efficient selection and handling features to improve accuracy. **Sushma S A** et al [13] proposed Comparative Study for prediction of heart diseases using three classifiers Naive Bayes, GNB and decision tree algorithms. an automated system was designed for efficient prediction which provides information for risks faced by the patients for heart diseases. **Jasna P. S.** et al [14] proposed a system named D-Predict. They explain that existing

systems are narrowed down for disease prediction when a few diseases or medical crucial. Proposed is based on input in the form of speech, using NLP. They extracted features and applied Naive Bayes classifier to predict disease the disease efficiently. **Daniel Jurafsky** et al [17] proposed Sentiment Classification using Naive Bayes. They applied Naive Bayes to text categorization; they assigned label or category to an entire document. They focused with only one common categorization task for sentiment analysis; they used ex-sentiment analysis for sentiment having positive or negative orientation expressed by writer for some object. **Chingmuan kim** et al [21] proposed Naive Bayes Classifiers with Text Classification". They applied textual features and represented them into vector form using extraction techniques of NLP. They applied Term Frequency and Bag of Words for computation. Their objective is to test imbalance dataset with Bayes Model with improved technique.

5. PROPOSED METHOD (GNB CLASSIFIER)

Following number of steps are used in gaussian bays classifiers.

Steps 1. Find 'Prior' probabilities for each given class.

Steps 2. Find probability of evidence that is used in the denominator.

Steps 3. Find the likelihood probability for evidence that is used in the numerator.

Steps 4. Substitute this probability into the Naive Bayes formula.

Steps 5. Check for usual assumption for attributes have normal or Gaussian distribution.

Steps 6. Select Probability density function (PDF) for the distribution.

Steps 7. Used probability density function

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-(x-\mu)^2/2\sigma^2}$$

For given x evaluate probability according to the distribution.

Steps 8. Calculate value for probability density function we need to calculate the following things

We approximate μ by the sample mean

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

We approximate σ^2 by the sample variance.

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n} (x_i - \bar{x})^2$$

6. ILLUSTRATE WITH AN EXAMPLE

Table 1 show sample data set which contains four attribute features with one outcome approved. Loan can be approved by using value of these four features. Features have categorical value.

Table 1: simple example of multi features dataset with categorial attribute

JobType Private	Income High	Credit Score High	Collateral F	Approved N
Private	High	High	T	N
Government	High	High	F	Y
SelfEmployed	Medium	High	F	Y
SelfEmployed	Medium	Medium	F	Y
SelfEmployed	Medium	Low	T	N
Government	Low	Low	T	Y
Private	Medium	High	F	N
Private	Medium	Low	F	Y
SelfEmployed	High	Medium	F	Y
Private	High	Low	T	Y
Government	Medium	High	T	Y
Government	High	Medium	F	Y
SelfEmployed	Medium	High	T	N

A new data came for prediction.

Table 2: new data for prediction

JobType	Income	CreditScore	Collateral	Approved
Private	Medium	High	T	?

Table 2 shows new tuple need to validate for loan approved.

P(Yes | E) = P(Private | yes) * P(Income | yes)

*P(CreditScore =High | yes) *P(Collateral =T | yes)
*P(Yes) / P(E)
= 0.0043 / P(E)
1/P(E) is normalization constant

P(no | E) = P (Private | no) *P(Income | no)
*P(CreditScore =High | no) *P(Collateral=T | no)
*P(no) / P(E)
= 0.0206 / P(E)
P(Approved =yes | E) + P(Approved =no | E)
= 0.0043 + 0.0206
P(Approved =yes | E) = 17.9%
P(Approved =no | E) = 82.1%

Data Set with numeric attribute data example

Table 3: Sample data set with continuous value

JobType	Income	CreditScore	Collateral	Approved
Private	86	85	F	N
Private	81	89	T	N
Government	82	87	F	Y
SelfEmployed	71	95	F	Y
SelfEmployed	67	81	F	Y
SelfEmployed	64	71	T	N
Government	65	66	T	Y
Private	73	94	F	N
Private	68	71	F	Y
SelfEmployed	74	79	F	Y
Private	74	71	T	Y
Government	73	91	T	Y
Government	82	76	F	Y
SelfEmployed	70	92	T	N

Table 3 shows the same data which used in table 1 but with original continues(numerical) value (not convert into categorical)

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-(x-\mu)^2/2\sigma^2}$$

Calculate the probability of Income =66 for class Yes:

μ (mean) 73

σ2 (variance) = 38

$$f(x/yes) = \frac{1}{\sqrt{2 * 3.14 * 38}} 2.7^{-(x-73)^2/2 * 38}$$

$$f(x/yes) = \frac{1}{\sqrt{2 * 3.14 * 38}} 2.7^{-(66-7\ )^2/2 * 38}$$

Table 4 shows the Computed probability of yes

Approved	Private	Medium	High	Collateral
Yes: 9	2/9	3/9	3/9	3/9
No: 5	3/5	1/5	4/5	3/5
Total	5	4	7	6

and no class for all attributes

P(Income =66|yes) =0.034

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-(x-\mu)^2/2\sigma^2}$$

Compute the probability of CreditScore =90 for class Yes:

μ (mean) = 79

σ2 (variance) = 104

$$f(x/yes) = \frac{1}{\sqrt{2 * 3.14 * 104}} 2.7^{-(x-79)^2/2 * 104}$$

$$f(x/yes) = \frac{1}{\sqrt{2 * 3.14 * 104}} 2.7^{-(90-7\ )^2/2 * 38}$$

P (CreditScore =90|yes) =0.022

P (Approved =yes | E) =

P (JobType =Private | Approved =yes) *P (Income =66 | Approved =yes) *P (CreditScore =90 | Approved =yes) *P (Collateral =True | Approved =yes) *P (Approved =yes) / P(E) = 0.000036 / P(E)

P (Approved =no | E) =

P (JobType =Private | Approved =no) *P (Income =66 | Approved=no) *P (CreditScore y=90 | Approved =no) *P (Collateral =True | Approved =no) *

P (Approved =no) / P(E) = 0.000136 / P(E)

After normalization: P(Approved =yes | E) = 19.9%, P(Approved =no | E) = 78.5%

7. IMPLEMENTATION

7.1 Implementation Environment

We evaluate the performance of both approach Categorial NB and Gaussian NB with the given dataset which have continue and discreate value. The experiments were performed on Intel Core i5 processor 8GB main memory Inbuilt hard disk 500GB and Windows7 OS. Python language is used to implement proposed approach using and dataset stored in CSV.

7.2 Description about the Dataset

Table 5 shows feature name and data types; total 19 attributes have been used, some of them are categorical and some of them are numerical. Data set contains mixed value. The Loan Approval Dataset contains data related to applicants applying for a home loan for HDFC bank. It is used to forecast whether a loan will be got approval (Y) or not (N) based on features of applicant.

The dataset combines both categorical and numerical features. No missing data ideal for direct model training. Several log-transformed columns are included to normalize skewed distributions (useful for models like Naive Bayes or logistic regression). The dataset is balanced enough for binary classification tasks.

Table 5 Compute the probability of yes and no

Journal of Theoretical and Applied Information Technology
15th February 2026. Vol.104. No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Feature Name	Data Type
1. Loan_No	Object
2. Gender	Categoric
3. Matrial status	Categoric
4. Dependents	Categoric
5. Qualification	Categoric
6. Self_Employed	Categoric
7. Applicant_Income	Numeric
8. Coapplicant_Income	Numeric
9. Loan_Amount	Numeric
10. Loan_Term	Numeric
11. Credit_record	Numeric
12. Property_Location	Categoric
13. Loan_Status	Category
14. Total_Income	Numeric
15. ApplicantIncomeLog	Numeric
16. CoapplicantIncomeLog	Numeric
17. LoanAmountLog	Numeric
18. Loan_Amount_Term_Log	Numeric
19. Total_Income_Log	Numeric

```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder
from sklearn.naive_bayes import CategoricalNB, GaussianNB
from sklearn.metrics import accuracy_score, classification_report, confusion_matrix
import matplotlib.pyplot as plt
import seaborn as sns
```

Figure. 1 Importing required library

```
# 2. Preprocess Data
# ----------------------------
df_clean = df.drop(columns=["Loan_ID"])  # Drop ID column
# Encode categorical columns
categorical_cols = df_clean.select_dtypes(include=["object"]).columns
label_encoders = {}
for col in categorical_cols:
    le = LabelEncoder()
    df_clean[col] = le.fit_transform(df_clean[col])
    label_encoders[col] = le
```

Figure. 2 Data Cleaning and label encoding

Dataset Overview

- Total Records (Rows): 6000
- Total Features (Columns): 19
- Target Variable: Loan_Status (Y = Approved, N = Not Approved)

```
# ----------------------------
# 3. Train Models
# ----------------------------
# CategoricalNB
cat_nb = CategoricalNB()
cat_nb.fit(X_train, y_train)
y_pred_cat = cat_nb.predict(X_test)

# GaussianNB
gauss_nb = GaussianNB()
gauss_nb.fit(X_train, y_train)
y_pred_gauss = gauss_nb.predict(X_test)
```

Fig. 3 Training Categorical NB and Gaussian NB

```
# ----------------------------
# 5. Plot Metrics Comparison
# ----------------------------
results = {
    "CategoricalNB": {
        "accuracy": acc_cat,
        "report": classification_report(y_test, y_pred_cat, output_dict=True)
    },
    "GaussianNB": {
        "accuracy": acc_gauss,
        "report": classification_report(y_test, y_pred_gauss, output_dict=True)
    }
}
```

Figure. 4 Testing Categorical NB and Gaussian NB

Figure 1 shows the library required for implementation. We import these libraries into our source code. Figure 2 shows code used for Data Cleaning and label encoding. Figure 3 shows the code for training Categorical NB and Gaussian NB. Figure 4 shows code for testing categorical

8. RESULT AND FINDING

8.1 Comparison of CategoricalNB vs GaussianNB

Some of the key factors need ti used to visually compares the performance Categorical NB and Gaussian NB these are :

1. Accuracy
2. Precision
3. Recall
4. F1-score

Each color bar represents a different metric:

1. Yellow: Accuracy
2. Orange: Precision
3. Red: Recall
4. Pink: F1-score

Table 6 Value for accuracy, precision, recall and F1 value

Naïve Bays	Accuracy	Precision	Recall	F1-score
CNB	0.6991	0.6650	0.6991	0.6187
GNB	0.8373	0.8384	0.8373	0.8275

Table 6 shows value of key factors after implementation needed to CategoricalNB and GaussianNB on same dataset. The x-axis shows the two models CategoricalNB and GaussianNB, while the y-axis shows the performance scores ranging from 0 to 1. Figure 6 shows accuracy value for CategoricalNB and GaussianNB, accuracy value of GaussianNB (0.837398) is higher as compared to CategoricalNB(0.699187).

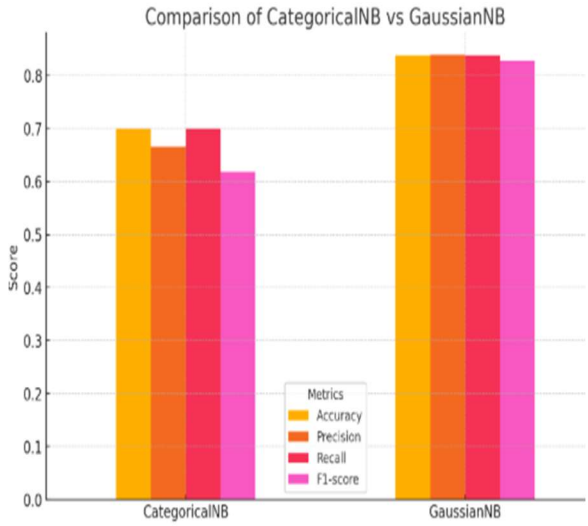

Figure 5: Graphical comparison of accuracy, precision, recall and F1 value

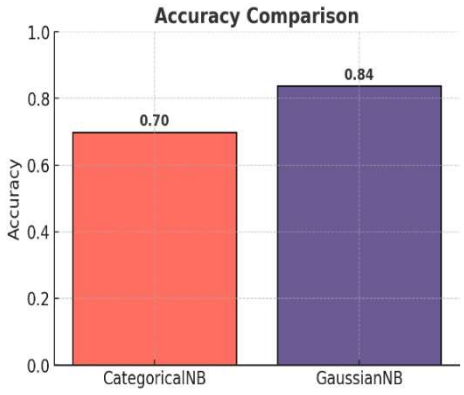

Figure 6 Comparing accuracy value

From the figure 7 we can see that precision value of GaussianNB (0.838482) is higher as compared to the precision value of CategoricalNB (0.665045). Figure 8 shows precision value for CategoricalNB and GaussianNB, recall value of GaussianNB (0.838482) is higher as compared to CategoricalNB (0.665045).

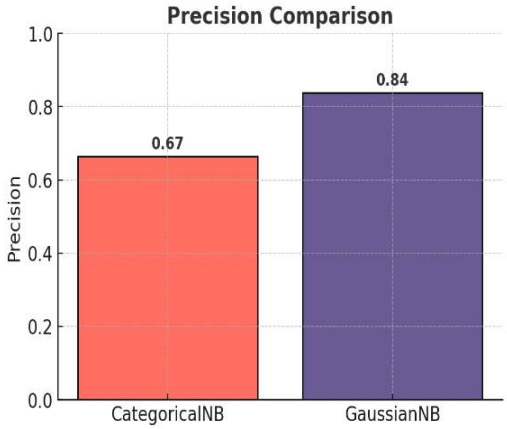

Figure 7 Comparing precision value

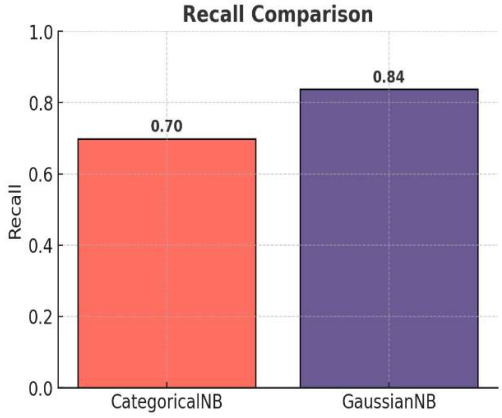

Figure 8 Comparing recall value

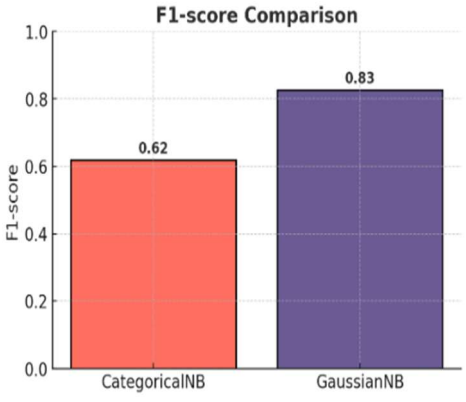

Figure 9 Comparing F1 value

Figure 9 shows F1-score value for CategoricalNB and GaussianNB, F1-score value of GaussianNB (0.827515) is higher as compared to CategoricalNB(0.8275157).

8.2 Comparing with confusion matrix for TP and FP.

1. Gaussian Naive Bayes (Right – Blue)
   TP (834): Correctly predicted Y
   TN (343): Correctly predicted N.
   FP (23): Rejected incorrectly predicted as Y.
   FN (0): Approved incorrectly predicted N.

2. Categorical Naive Bayes (Left – Green)
   TP (779): Approved correctly Y.
   TN (166): Rejected correctly N.
   FP (200): Rejected misclassified Y.
   FP (55): Approved misclassified as N.

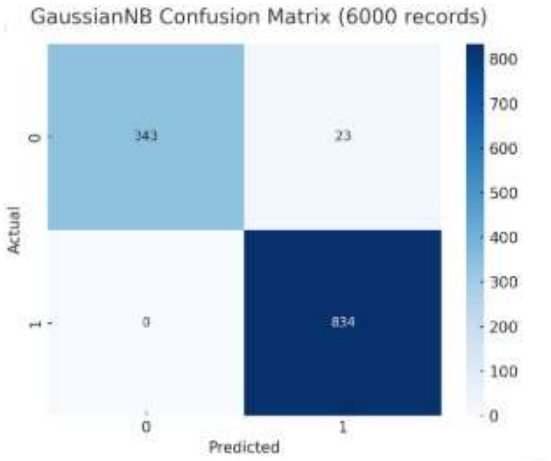

Figure 10 Confusion matrix for TP and FP using GaussianNB

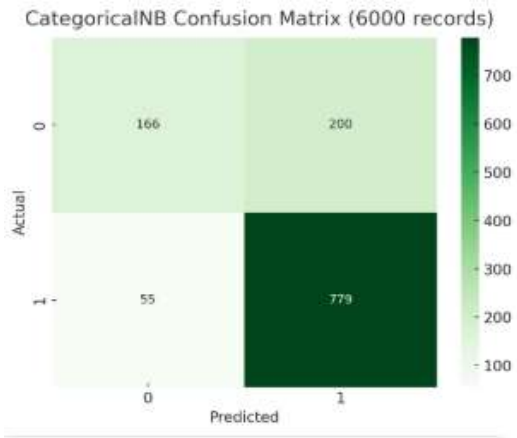

Figure 11Confusion matrix for TP and FP using CategoricalNB

In testing we can see from figure 10 and 11 it is found number of records that are correctly classified by the GaussianNB is much higher than as compared to CategoricalNB.

9. CONCLUSION AND FUTURE WORK

In this study, Categorical Naïve Bayes (CategoricalNB) and Gaussian Naïve Bayes (GaussianNB) classifiers were applied to a real-life dataset obtained from the HDFC Bank loan application department, consisting of 6,000 records with both categorical and continuous attributes. The CategoricalNB model showed reasonable performance; however, it exhibited higher class confusion due to the required discretization of continuous attributes. This discretization process led to a greater number of misclassifications, indicating that the categorical assumption does not fully align with the dataset's characteristics.

In contrast, the GaussianNB model achieved higher accuracy and consistency, as it can directly handle continuous attributes without discretization. It demonstrated improved class separation between approved and non-approved loan applications, resulting in fewer misclassifications and higher precision. Although the CategoricalNB classifier remains acceptable for datasets dominated by categorical features, it was found to be less stable for datasets containing mixed attribute types.

Overall, the experimental results confirm that Gaussian Naïve Bayes performs more effectively than Categorical Naïve Bayes for datasets with continuous-valued attributes.

10. LIMITATIONS AND FUTURE SCOPE

Zero probability is difficult to handle for both methods and not difficult to predict. In both cases implicitly assume predictors it rarely happens in real life.
In future we will work for hybrid models for classification using on the data set which have mixed attribute categorical and numerical In future we will use other classifiers and compare accuracy with hybrid model.

REFERENCES

[1]. R. Nithya Dr. D. Ramyachitra P. Manikandan "An Efficient Bayes Classifiers Algorithm on 10-fold Cross Validation for Heart Disease Dataset" International Journal of Computational Intelligence and Informatics, Vol. 5: No. 3, December 2015.

[2]. Ruth Talbot, Chloe Acheampong and Richard Wicentowski "SWASH: A Naive Bayes Classifier for Tweet Sentiment Identification" Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), pages 626–630, Denver, Colorado, June 4-5, 2015.

[3]. Unggul Widodo Wijayanto Riyanarto Sarno An Experimental Study of Supervised Sentiment Analysis Using Gaussian Naïve Bayes 2018 International Seminar on Application for Technology of Information and Communication (iSemantic)

[4]. Nafizatus Salmi and Zuherman Rustam Naïve Bayes Classifier Models for Predicting the Colon Cancer 9th Annual Basic Science International Conference 2019 IOP Conf. Series: Materials Science and Engineering 546 (2019)

[5]. Allsela Meiriza , Endang Lestari, Pacu Putra, Ayu Monaputri , and Dini Ayu Lestari Prediction Graduate Student Use Naive Bayes Classifier Advances in Intelligent Systems Research, volume 172 Sriwijaya International Conference on Information Technology and Its Applications (SICONIAN 2019)

[6]. Shikha Agarwal, Balmukumd Jha, Tisu Kumar, Manish Kumar, Prabhat Ranjan Hybrid of Naive Bayes and Gaussian Naive Bayes for Classification: A Map Reduce Approach International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8, Issue-6S3, April 2019

[7]. Amitha, Merin Meleet A System for Recommendation of Medication Using Gaussian Naïve Bayes Classifier International Journal of Innovative Research in Computer Science & Technology ISSN: 2347-5552, Volume-7, Issue-3, May-2019

[8]. Marzuki Ismail, Norlida Hassan, Salem Saleh Bafjaish Comparative Analysis of Naive Bayesian Techniques in Health-Related for Classification Task Journal of Soft Computing and Data Mining Vol.1 No. 2 (2020) 1-10 ISSN: 2716-621X

[9]. Jiajie Shen and Hongqing Fang Human Activity Recognition Using Gaussian Naïve Bayes Algorithm in Smart Home AICS 2020 Journal of Physics: Conference Series 1631 (2020) 012059 IOP Publishing doi:10.1088/1742-6596/1631/1/012059

[10]. Brijmohan Lal Sahu Dr. Anil Tiwari Student placement possibility prediction using Naive Bayes algorithm International Journal of Advance Research, Ideas and Innovations in Technology ISSN: 2454-132X Impact factor: 6.078 (Volume 6, Issue 3) Available online at: www.ijariit.com

[11]. Krish Shah, Rajiv Punjabi, Priyanshi Shah Real Time Diabetes Prediction using Naïve Bayes Classifier on Big Data of Healthcare International Research Journal of Engineering and Technology ISSN: 2395-0056 Volume: 07 Issue: 05 | May 2020

[12]. Hong Chen , Songhua Hu , Rui Hua and Xiuju Zhao Chen et al. EURASIP Journal on Advances in Signal Processing (2021) 2021:30

[13]. Sushma S A , Keerthan Kumar T G Comparative Study of Naive Bayes, Gaussian Naive Bayes Classifier and Decision Tree Algorithms for Prediction of Heart Diseases International Journal for Research in Applied Science & Engineering Technology ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429 Volume 9 Issue

[14]. Jasna P. S., Aiswarya M. K., Krishnapriya K.V., Sisira P.C 5 D-Predict: Disease Prediction Using Gaussian Naive Bayes Algorithm ISSN (PRINT): 2393-8374, (ONLINE): 2394-0697, Volume-8, ISSUE-6, 2021 International Journal Of Current Engineering and Scientific Research.

[15]. M. Vijay Anand, B. KiranBala, S. R. Srividhya , Kavitha C., Mohammed Younus, and Md Habibur Rahman Gaussian Nave Bayes Algorithm: A Reliable Technique Involved in the Assortment of the Segregation in Cancer Hindawi Mobile Information Systems Volume 2022.

[16]. Naulak, Chingmuankim A comparative study of Naive Bayes Classifiers with improved technique on Text Classificationhttps://www.techrxiv.org 10.36227/techrxiv. 19918360.v1.

[17]. Daniel Jurafsky & James H. Martin. Naive Bayes and Sentiment Classification Speech and Language Processing. Daniel Jurafsky & James H. Martin. Copyright © 2023. All rights reserved. Draft of January 7, 2023.

[18]. Kalakonda Shashank, Anumandla Sahithya, Shaik Shakeel, Reviews analysis Using gaussian naïve bayesin machine Learning International Research Journal of Engineering and Technology ISSN: 2395-0056 Volume: 10 Issue: 01 | Jan 2023 www.irjet.net p-ISSN: 2395-0072

[19]. M. Vedaraj, C.S. Anita, A. Muralidhar, V. Lavanya, K. Balasaranya, Early Prediction of Lung Cancer Using Gaussian Naive Bayes

Classification Algorithm International Journal of Intelligent Systems A and Applications in Engineering ISSN:2147-67992147-6799 www.ijisae.org

[20]. Mokhairi Makhtar, Hasnah Nawang, Syadiah Nor Analysis On Students Performance Using Naïve Bayes Classifier Journal of Theoretical and Applied Information Technology 31st August 2017. Vol.95. No.16 2005

[21]. Chingmuankim and Prof.Rajni Jindal A comparative study of Naive Bayes Classifiers with improved technique on Text Classification Text Classification. TechRxiv. Preprint. https://www.techrxiv.org 28-05-2022 / 31-05-2022

[22]. B. M. Gayathr C. P. Sumathi, PhD an Automated Technique using Gaussian Naïve Bayes Classifier to Classify Breast Cancer International Journal of Computer Applications (0975 – 8887) Volume 148 – No.6, August 2016