Comparative Analysis of Naive Bayesian Techniques in Health-Related for Classification Task

Marzuki Ismail[1], Norlida Hassan[1*], Salem Saleh Bafjaish[2]

[1]Faculty of Computer Science & Information Technology,
 University Tun Hussein Onn Malaysia, Johor, 86400, MALAYSIA

[2]Department of Computer Science, Faculty of Oil and Minerals,
 University of Aden, Shabwah, 401, YEMEN

*Corresponding Author

Abstract: Naïve Bayes is a technique of using algorithms based on the Naïve Bayes theorem, which utilizes naive assumptions of conditional independence among predictors to predict the class of unknown data sets. The problems that face classification techniques are the accuracy of the classification and the number of errors classifying. However, it is been utilized as a classification that involves a several models for data mining in health like, Naïve Bays classifier which is used for the purpose of identifying the positive and the negative sentiments of the patients. Moreover, it's been used also integrated with machine learning for the purpose of opinion mining and sentiment classification as well as it utilized as a method for predicting the diseases. This paper aims for exploring the several different techniques that will give different results based on their respective algorithms. This research will focus on the comparative analysis of the differences in performance and type of variations of the Naïve Bayes classification. There are generally four applications that use Naïve Bayes, real-time prediction, multiclass prediction, text classification, and recommendation system. To overcome the drawbacks of these issues, this research will apply three Naive Bayes models namely as Gaussian, Multinomial, and Bernoulli model. These models fall under the same type of classification technique which uses the Bayes theorem. The Gaussian model is used in basic classification and assumes that the features of a dataset follow a normal distribution. The multinomial model, however, is used for discrete counts, such as counting how many numbers of times the outcome of x is observed over n number of trials. The Bernoulli model primarily focuses on searching for vector features that are binary. The objective is as follows, to apply and implement the original model Naïve Bayes with different existing models such as the Multinomial Naïve Bayes and the Gaussian, and the Bernoulli Naïve Bayes. The outcome of this study will focus on the differences, capabilities, and performance of the probabilistic classifier of the Naïve Bayes algorithms.

Keywords: Naïve Bayes, algorithms, data mining, classification

1. Introduction

Naïve Bayes algorithm is a classification technique based on the Bayes' Theorem with an assumption of independence among its predictors. In simple terms, a Naïve Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other features that are inherently codependent on one another.

Naïve Bayes model is easy to build and is particularly useful for very large data sets as it can be categorized as one of the fastest in terms of completion because of its simplicity.

Along with simplicity, Naïve Bayes is known to outperform even more highly sophisticated classification methods, such as stated by [1]. The use of data mining can play an important role in the enhancement and efficiency of healthcare systems. Different methods are related to the analysis of the diseases as stated in [2], as in their paper they state different method for the analysis of the data like the convolutional neural network in which it considered as a method for classification of the heartbeats using ECG signals, the support vector machine for the tooth detection images, breast cancer classification using naïve Bayes (NB) classifier. However, Naïve Bayes (NB) classifier is one of the supervised learning categories. Naive Bayes classifier assumes that the presence or absence of a particular feature of a class is unrelated to the presence or absence of other features. This means that the features selection of NB is mutually exclusive towards one another. This could be seen as a bonus or weakness depending on the data type and how it works.

The Naive Bayes classifier [3] was adopted because of its computational efficiency as well as its optimality for classification tasks even when the conditional independence assumption is invalid[4]. In [5], the public health takes much consideration for the public with the use of data mining; the use of data mining in the healthcare systems is been critical for many aspects like the personalization, the studies of the medical data and many additional aspects. Based on[5]taking the advantages of applying the Bayes' Theorem and prevalence statistics, dubbed naive Bayes classifiers, aim to accomplish this with readily available data.

The objectives of this study are to determine the best classification algorithm using different classification techniques incorporated with different variations of the algorithm. Therefore, this study has several objectives as follows:

1   To apply and implement the original model Naïve Bayes with different existing models such as the Multinomial Naïve Bayes and the Gaussian, and the Bernoulli Naïve Bayes using Scikit Learn (python library).
2   To simulate the proposed framework in the (1) for the classification task.
3   To evaluate the performance of the proposed simulation in (2) and benchmark the results with the (Naïve Bayes model).

The scope of study in this research is focusing on the different types of variations of the same three classification technique algorithms in machine learning, which are; Naïve Bayes, Multinomial Naïve Bayes, and Bernoulli Naïve Bayes. The study aims to identify the highest accuracy for classifying the different Naïve Bayes model using the different variations, which uses different mathematical algorithms while comparing it to the original model of Naïve Bayes.

The expected outcome of the research will focus on three main purposes. These objectives are reached where:

- Results of the comparison of the accuracy between different technique that can classify:
  a. Gaussian Naïve Bayes
  b. Multinomial Naïve Bayes
  c. Bernoulli Naïve Bayes
- Identify which is the best algorithm to classify with the highest accuracy using min-max normalization and true-positives and false-positives.
- Evaluate and analyze sources of information to discover through data mining techniques, to build data-driven models and extract useful knowledge.

2.   Related Works

In the terminology of machine learning, classification is considered an instance of supervised learning, i.e., learning where a training set of correctly identified observations is available. A study done by[6]shows the Naive Bayes model being compared between two variations, which was the Multinomial and Bernoulli Naive Bayes. These different variations are how the Naive Bayes will calculate the normal distributions based on their different algorithms. The Naive Bayes classifier falls into the probabilistic classifier category and is based on applying the Bayes Theorem with strong independence between assumptions of the features, hence the name of naive. A probabilistic classifier is a classifier that can predict, given an observation of an input, a probability distribution over a set of classes, rather than only outputting the most likely class that the observation should belong to.

A study done by[7] found that the probabilistic classifier, Naive Bayes is providing improved accuracy with low computational effort and very high speed. In the following subsections, literature is been explained to fully understand the concept of machine learning and naïve Bayes classifier. In [8] Gaussian Naïve Bayes algorithm is been implemented for classifying cancer using two datasets namely WBCD and lung cancer dataset.

2.1 **Classification in Machine Learning**

In machine learning, classification is a problem in which the machine has to identify a set of categories that are based on a training set of data containing attributes or instances whose category membership is known. An example of this is given a data set of health-related, either the patient has the disease or not. Classification is also an example of pattern recognition. The naive Bayes classifier is a collection of classification algorithms based on the Bayes' theorem. It is not a single algorithm but a family of algorithms where all of them share a common principle. The problem of this classifier can potentially be improved by using an ensemble to combine several classifiers and produce a better predictive performance. In [9]framework for the transfusion of the best CP integrated with machine learning is been used.

2.2 Machine Learning Classification Algorithm

Machine learning is defined as a set of algorithms or steps that can detect patterns, uncover patterns and even predict patterns of future automatically in data or performing different kinds of decision making under uncertainty [10].[11]SVM (linear) classifier is been used as the best diagnosis model for COVID19 as well as in [12]. A Naive Bayes classifier depicted as a Bayesian network in which the predictive attributes $x_1, x_2, \ldots x_k$ are conditionally independent given the class attribute.

Naive Bayes methods are a set of supervised learning algorithms based on applying Bayes' theorem with the "Naive" assumption of conditional independence between every pair of features given the value of the class variable. Bayes' theorem states the following relationship, given class variable y and dependent feature vector x1 through x_n,:

$$P(y \mid x_1, \ldots, x_n) = \frac{P(y)P(x_1, \ldots x_n \mid y)}{P(x_1, \ldots, x_n)} \qquad (1)$$

Using the Naive conditional independence assumption that

$$P(x_i \mid y, x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) = P(x_i \mid y), \quad (2)$$

For all i, this relationship is simplified to

$$P(y \mid x_1, \ldots, x_n) = \frac{P(y)\prod_{i=1}^{n} P(x_i \mid y)}{P(x_1, \ldots, x_n)} \qquad (3)$$

Since $P(x_1, \ldots, x_n)$ is constant given the input, we can use the following classification rule:

$$P(y \mid x_1, \ldots, x_n) \propto P(y)\prod_{i=1}^{n} P(x_i \mid y) \qquad (4)$$

$$\Downarrow$$

$$\hat{y} = \arg \max_{y} P(y)\prod_{i=1}^{n} P(x_i \mid y), \qquad (5)$$

In addition, we can use Maximum A Posteriori (MAP) estimation to estimate $P(y)$ and P(xi | y); the former is then the relative frequency of a class y in the training set. According to[13] the different Naive Bayes classifiers differ mainly by the assumptions they make regarding the distribution of P(xi | y).

2.3 Naive Bayes Classifier

The Naive Bayes classifier provides a simple approach with clear semantics to represent the learning probabilistic knowledge of the Bayes theorem. Another assumption many view the classifier as a form of Bayesian network that is termed naive because it relies on two important simplifying assumptions. In particular, it assumes the predictive attributes are conditionally independent given the class and it shows no hidden or latent attributes that can influence the prediction process. A graphical representation of the Naive Bayesian classifier follows that depicted in Fig. 1 below:
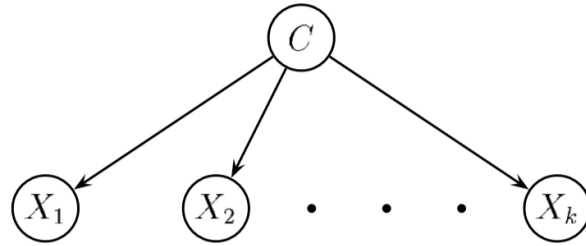
Fig. 1 - The graphical model of Naive Bayes [13]

In [14], there was an investigation carried out for the purpose of evaluation of the performance of the machine learning tool. However, the new naïve Bayes classifier can be used with the weight method in order to classify breast cancer. Moreover, the new tool is been used to enhance and improve the accuracy of breast cancer detection in the field of medical data mining. Additionally, the new tool has promised with strong accuracy in terms of the compression with other classifiers as the following Table 1. In [8] classification that is implemented with the use of naïve Bayes considered to be one of the best solutions when it comes to the health systems because it is the simplest form of Bayesian network classifier based on applying Bayes theorem, with strong independence of attributes assumption.

Table 1 - The methodology Phases[14]

S, no	Data set	Classifiers	Percentage
1	WBC	Weighted associated classifier	90.41%
2	WBC	Fuzzy associated classifier	95.10%
3	WBC	CBA	93.79%
4	WBC	CMAR	88.812%
5	WBC	CPAR	92.84%
4	Large data set	Radial basis	87.42%
5	Large data set	Decision tree	85.71%
6	Large data set	Nearest neighbor	84.57%

A. Gaussian Naive Bayes Classifier

The Gaussian Naive Bayes is a variant of Naive Bayes that implements the Gaussian normal distribution and supports continuous data for classification. The likelihood of the features used is assumed to be using Gaussian:

$$P(x_i \mid y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left(-\frac{(x_i - \mu_y)^2}{2\sigma_y^2}\right) \quad (6)$$

By using this formula, we can calculate the probability of the classification data to fall within the normal distribution of the Gaussian algorithm.

B. Multinomial Naive Bayes Classifier

The Multinomial Naive Bayes algorithm implements the Naive Bayes algorithm for multinomial distributed data and is one of the two classic Naive Bayes variants used in text classification[15]. The data are typically represented as word vector counts. The distribution of word vector count is parametrized by vectors $\theta_y = (\theta_{y1}, \ldots, \theta_{yn})$ for each class y, where n is the number of features (in text classification, the size of the vocabulary) and θ_{yi} is the probability $P(x_i \mid y)$ of feature i appearing in a sample belonging to class y.

The parameter θ_y is estimated by a smoothed version of maximum likelihood, i.e. relative frequency counting:

$$\hat{\theta}_{yi} = \frac{N_{yi} + \alpha}{N_y + \alpha n} \quad (7)$$

Where $N_{yi} = \sum_{x \in T} x_i$ is the number of times feature i appear in a sample of class y in the training set, and $N_y = \sum_{i=1}^{n} N_{yi}$ is the total count of all features for class y. The smoothing priors $\alpha \geq 0$ account for features not present in

the learning samples and prevent zero probabilities in further computations. Setting $\alpha = 1$ is called Laplace smoothing, while $\alpha < 1$ is called Lidstone smoothing[16].

C. Bernoulli Naive Bayes Classifier

In the multivariate Bernoulli Naïve Bayes Classifier algorithm, features are independent binary variables, which represents that whether a term is present in the document under consideration, or not[6]. The decision rule for Bernoulli Naive Bayes is based on:

$$P(x_i \mid y) = P(i \mid y)x_i + \big(1 - P(i \mid y)\big)(1 - x_i) \qquad (8)$$

Multi-variate Bernoulli performs well with small vocabulary but that the multinomial performs usually performs even better at larger vocabulary sizes--providing on average a 27% reduction in error over the multi-variate Bernoulli model at any vocabulary size [16]. Bernoulli Naive Bayes might perform better on some datasets, especially those with shorter documents.

3. Methodology

The methodology part for this research uses the CRISP-DM [17] method to critically analyze the processes that are important in producing viable and feasible data mining research. By using the five processes in CRISP-DM methodology, the paper will introduce the adoption and implementation of the various classification models while applying the various techniques that were stated in section 2. Fig. 2 illustrates the graphical model for building a data mining technique. Firstly, understating the business in which it leads to fully understand how the comparative analysis will be carried out and it will benefit the next step in terms of the amount of data that will be utilized to achieve the best result of the comparative analysis. Secondly, data used has to be identified and classifies the sort of it which leads to better performance and result. Thirdly, data preparation data used has to be filtered first in which it leads to better performance and result. Fourthly, preparing the models that will be used in the comparative analysis as they are namely BernoulliNB, MultinomialNB and GaussianNB. Fifthly, the evaluation of the models plays important role in achieving a better result. Lastly, the development phase is done based on the previous phases with the use of machine learning concept.
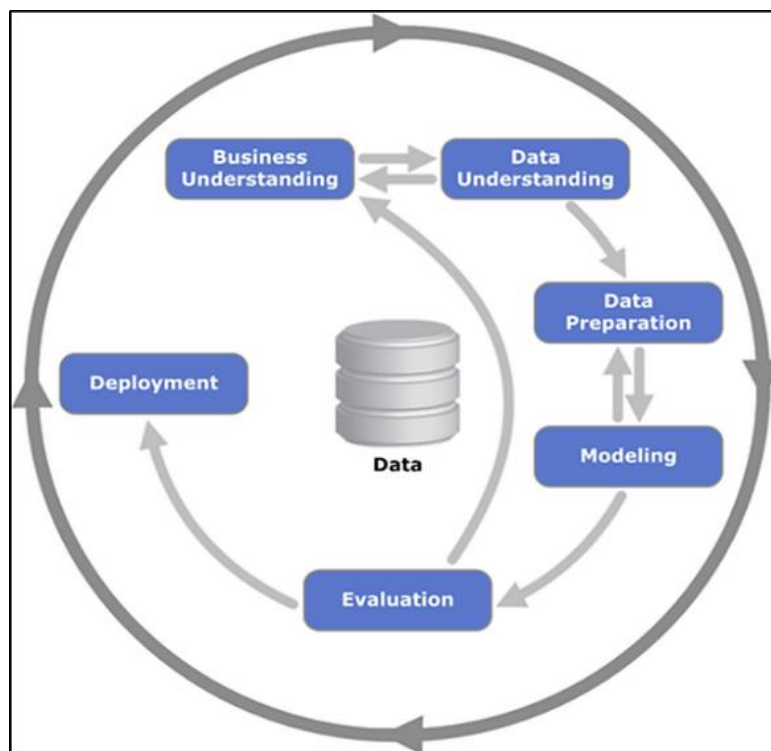

Fig.2 - Framework for comparative analysis between Naive Bayes Techniques in health-related for classification task [17]

The research framework is based on the methodology and implements the phases of the CRISP-DM. The steps are shown below in Table 2 and Fig. 3 on how to produce the results for this project.

Table 2 - The methodology Phases

| Phase	Description/Explanation
Data Understanding	The finding of the data, where it was obtained, introduces the type of dataset, what type of attributes and number
Data Preparation	Preparation of the dataset, replacing missing values, data preprocessing, normalization of data, conversion of data
Model Building	Preparing the model by using multiple classification techniques chosen, introduces other techniques
Evaluation	Evaluates the results based on data analysis, uses common evaluation techniques, accuracy, precision, and recall
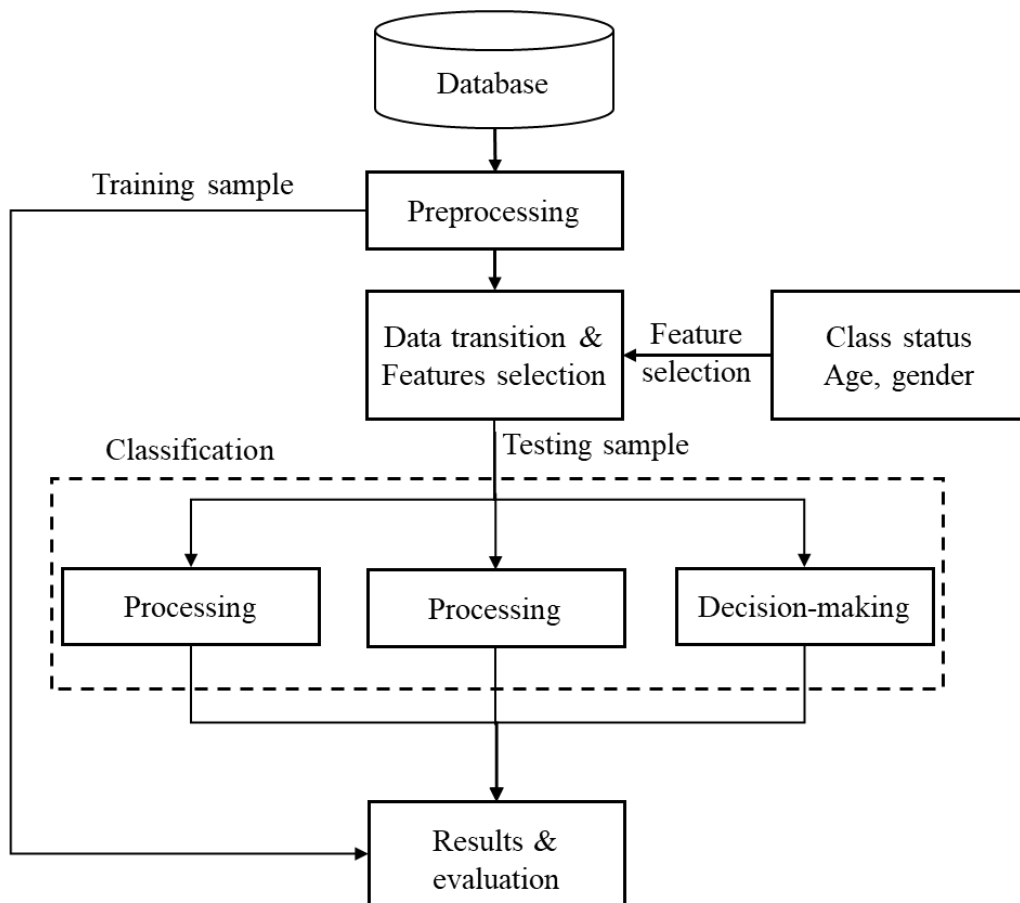
Fig. 3 - Framework for creating this project based on the classification models

3.1 Data Understanding Phase

The dataset was collected from various websites that had an abundance of data that was collected and stored. These websites are renowned for their officially cited dataset that was gathered by researchers throughout the years. The website that was gathered from UCI. The data used was health-based, and was suitable for supervised learning classification. The data that was taken involved the classification of breast cancer (benign or malignant) and heart disease. It is used to determine whether the patient suffered from these conditions.

3.2 Data Preparation

The dataset was replaced with all missing attributes and values. The dataset also went through the cleansing phase. The dataset was then preprocessed through normalization. Finally, it was parsed through again to make sure that all the classification techniques could handle polynomial to the numerical conversion of the data.

3.3 Classification Algorithm

The chosen machine learning algorithm for this research project is the Naive Bayes algorithm. In it, there are multiple variants of the same algorithm that is different in their calculations and classifications method. In the Naive Bayes algorithm, there are two classes that dictate what the outcome of the result will be. The two classes are labelled as $i = (0,1)$. In this instance, the classes labelled is to construct a score set that is associated with class 1 and class 0 objects. The Naive Bayes algorithm will classify which class it belongs to based on the given dataset as shown in Fig 4.

```
1. for q = 1 ... s // loop for each mining model's element
2.     μ[q] = 0; // initialization of mining model's elements
3. for j = 1 ... μ  // loop for each vector
4.     μ[d[j,p]]++; //increment count of vectors for value x_{j,p} of vector x_j;
5.     for k = 1 ...p-1 // loop for each attribute
6.         μ[φ(k-1)+(d[j, k]-1)·φ(0)+ d[j, p]]++; // increment count of vectors with
                                                  // value x_{j,k}  and value x_{j,p}
7.     end for;
8. end for;
```

Fig. 4 - Naive Bayes algorithm

3.4 Model Building

Model building is the phase which to train the dataset using the selected algorithm that was stated earlier. This project uses Scikit-Learn software to embed the Naive Bayesian classifier into the code. Scikit Learn is a software that uses Python Programming Language to create classification models. It is to also create the training and testing epochs that can be seen as a test for the machine learning abilities.

Naive Bayes could be a straightforward probabilistic classifier that calculates a collection of chances by forwarding the frequency and combos of values from the given datasets. The algorithm uses the Bayes theorem and assumes all the independent or non-interdependent attributes given by the value of the class variable. Naive Bayes is based on a simplified assumption that attribute values are conditional on each other free of charge if given output value. In other words, given the output value, the probability of collectively observing is the product of the individual probability. Besides that, the Bayes algorithm is based on posterior probability, $P(c|x)$ from(c), $P(x)$and $P(x|c)$. Naive Bayes classifier assumes that the effect of the value of a predictor (x) on given class (c) is independent values of other predictors. According to Bayes theorem, the equation below shows how to calculate class independence.

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)} \quad (9)$$

where $P(c|x)$ is the posterior probability of class (target) given predict (attribute), $P(c)$ is the prior probability of a class, $P(x|c)$ is the likelihood which is the probability of predictor given class and $P(x)$ is the prior probability of predictor.

The Multinomial, Bernoulli, and Gaussian Naïve Bayes classifiers have been generated using the following code of Fig 5 on the training set [14].

7

```
from sklearn.datasets import load_breast_cancer
from sklearn.naive_bayes import MultinomialNB, BernoulliNB, GaussianNB
mnb = MultinomialNB()
mnb.fit()
mnb.score()
bnb = BernoulliNB()
bnb.fit()
bnb.score()
gnb = GaussianNB()
gnb.fit()
gnb.score()
```

Fig. 5 - Code for building the NB classifiers in python

4.   Research Design and Implementation

In this paper, the main measure of performance is evaluated in terms of accuracy, precision, and recall from the confusion matrix of classification. The measures are computed by using equations that are described below:

- **Accuracy:** It is the total number of samples correctly classified to the total number of samples classified. The formula for calculating accuracy is shown in Equation 1.

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + TN + FP + FN)} \qquad (10)$$

  where TP is True Positive, TN is True Negative and FN is False Negative.

- **Precision:** It is the number of samples is categorized positively classed correctly divided by total samples are classified as positive samples. The formula for calculating precision is shown in Equation 2.

$$\text{Precision} = \frac{TP}{(TP + FP)} \qquad (11)$$

  where TP is True Positive and FP is False Positive.

- **Recall:** It is the number of samples is classified as positive divided by the total sample in the testing set positive category. The formula for calculating recall is shown in Equation 3.

$$\text{Recall} = \frac{TP}{(TP + FN)} \qquad (12)$$

Below is the algorithm that was used.

Sklearn Gaussian algorithm
cl_gauss = **sklearn**.naive_bayes.**GaussianNB**()
res_gauss = cl_gauss.fit(X_train_voc, y_train).predict(X_test_voc)
metrics.accuracy_score(y_test, res_gauss) * 100

Sklearn Multinomial algorithm:
cl_multi = **sklearn**.naive_bayes.**MultinomialNB**()
res_multi = cl_multi.fit(X_train_voc, y_train).predict(X_test_voc)
metrics.accuracy_score(y_test, res_multi) * 100

Sklearn Bernoulli algorithm:
cl_bern = **sklearn**.naive_bayes.**BernoulliNB**()
res_bern = cl_bern.fit(X_train_voc, y_train).predict(X_test_voc)
metrics.accuracy_score(y_test, res_bern) * 100

5.   Result and Discussions

To investigate the accuracy of the three classification models, two benchmark datasets were used. These datasets are Breast Cancer, and Heart Disease has taken from the UCI machine learning repository. Naive Bayes performance will be indicated by using the different variations of calculating its normal distribution. Therefore, this experiment will focus on the accuracy of the variations. Table 3 summarizes the characteristics of the datasets used in the experiments.

Table 3 - Characteristics of datasets

Dataset	Examples	Train data	Class	No. Instance in each class	No. of features
Breast cancer	569	171	2	C1=86, C2=85	30
Heart Disease	300	90	2	C1=30, C2=30	14

For training of the algorithms, Table 3 mentions the features that were tested upon. The first dataset was trained initially using 30 features and then 15 and lastly 5 features to achieve the mean output of the testing. The second dataset was trained initially using 14 features and lastly 5 features. For training the three classification models, Table 4 splits the data. In this section, we report the results in the form of accuracy, mean accuracy (from both datasets) and TPR value.

Table 4 - Train and test split dataset

Method	Dataset 1 (Breast Cancer)		Dataset 2 (Heart Disease)	
	Train	Test	Train	Test
BernoulliNB	70%	30%	70%	30%
MultinomialNB	70%	30%	70%	30%
GaussianNB	70%	30%	70%	30%

Table 5 shows the datasets that are split into training and testing. We have chosen 70:30 ratio because the Naive Bayes classifier is simplistic and cannot benefit from overfitting the data.

Table 5 - Experimental results of classification

	Accuracy%		Mean Accuracy%	TPR (True Pos. Rate)	
	Breast Cancer	Heart Disease		Breast Cancer	Heart Disease
BernoulliNB	96.90	96.42	96.66	0.887	0.884
MultinomialNB	97.12	97.02	97.07	0.973	0.972
GaussianNB	96.98	96.79	96.89	0.955	0.950

Table 5 documents the results of classification, accuracy and TPR of the different models Naive Bayes. In this study, Multinomial Naive Bayes achieved the highest mean accuracy with 97.07%. This shows that the Multinomial outperformed the other Bernoulli and Gaussian models. Fig 6 shows the comparison of accuracies of three depicted by a Bar graph
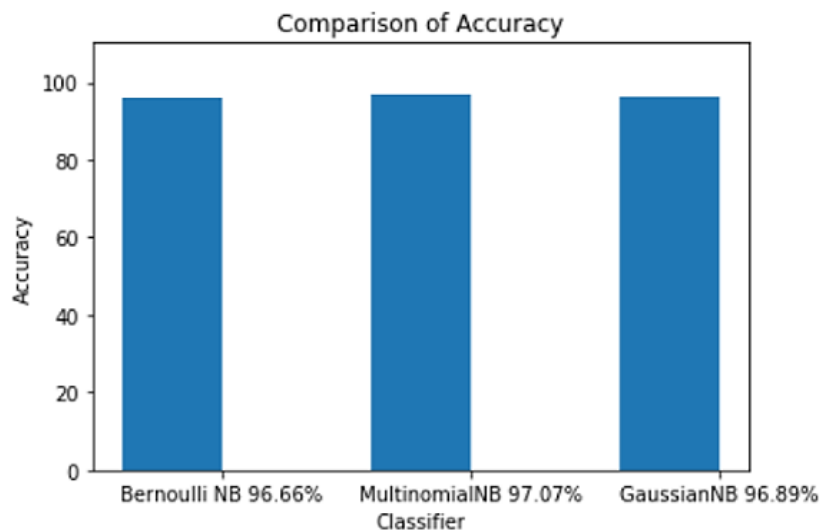
Fig. 6 - Comparison of accuracies of three depicted by a bar graph

6.   Conclusion

In this paper, we used three algorithms to objectively show which algorithm is better at classifying these datasets. The parameters that were involved were the feature selection and removal. We tested with full features and then gradually decreased the set of features to deduce which algorithm is better at classifying with less and fewer features. The simulation results show that the Multinomial Naive Bayes has better accuracy and mean accuracy when compared with the other two techniques given the same dataset and parameters. In future work, there will be two aspects have to be taken into consideration, namely, more algorithms can be compared to achieve better results and potentially introduce a better algorithm in Naive Bayes. Moreover, the comparison will be carried out with new naïve algorithms to evaluate the performance of them to justify to be used among the health systems.

Acknowledgement

References

[1]   Ashari, Ahmad, Iman Paryudi, and A. Min Tjoa. "Performance comparison between Naïve Bayes, decision tree and k-nearest neighbor in searching alternative design in an energy simulation tool." *International Journal of Advanced Computer Science and Applications (IJACSA)* 4.11 (2013).

[2]   Al-Aidaroos, K. M., Bakar, A. A., & Othman, Z. (2012). Medical data classification with Naive Bayes approach. *Information Technology Journal*, *11*(9), 1166.

[3]   Langley, P., Iba, W., & Thompson, K. (1992, July). An analysis of Bayesian classifiers. In *Aaai* (Vol. 90, pp. 223-228).

[4]   Domingos, P., & Pazzani, M. (1997). On the optimality of the simple Bayesian classifier under zero-one loss. *Machine learning*, *29*(2-3), 103-130.

[5]   Hickey, S. J. (2013). Naive Bayes classification of public health data with greedy feature selection. *Communications of the IIMA*, *13*(2), 7.

[6]   G. Singh, B. Kumar, L. Gaur, and A. Tyagi, "Comparison between Multinomial and Bernoulli Naïve Bayes for Text Classification," *2019 Int. Conf. Autom. Comput. Technol. Manag. ICACTM 2019*, no. May 2020, pp. 593–596, 2019, doi: 10.1109/ICACTM.2019.8776800.

[7]   Kharya, S., Agrawal, S., & Soni, S. (2014). Naive Bayes classifiers: a probabilistic detection model for breast cancer. *International Journal of Computer Applications*, *92*(10), 0975-8887.

[8]   Jabbar, M. A., & Samreen, S. (2016, October). Heart disease prediction system based on hidden naïve bayes classifier. In *2016 International Conference on Circuits, Controls, Communications and Computing (I4C)* (pp. 1-5). IEEE.

[9]   Albahri, O. S., Al-Obaidi, J. R., Zaidan, A. A., Albahri, A. S., Zaidan, B. B., Salih, M. M., ... & Aleesa, A. M. (2020). Helping doctors hasten COVID-19 treatment: Towards a rescue framework for the transfusion of best convalescent plasma to the most critical patients based on biological requirements via ml and novel MCDM methods. *Computer methods and programs in biomedicine*, *196*, 105617.

[10]   Tan, A. H. (1999, April). Text mining: The state of the art and the challenges. In *Proceedings of the pakdd 1999 workshop on knowledge disocovery from advanced databases* (Vol. 8, pp. 65-70). sn.

[11]   Mohammed, M. A., Abdulkareem, K. H., Al-Waisy, A. S., Mostafa, S. A., Al-Fahdawi, S., Dinar, A. M., ... & Arbaiy, N. (2020). Benchmarking Methodology for Selection of Optimal COVID-19 Diagnostic Model Based on Entropy and TOPSIS Methods. *IEEE Access*.

[12]   Mohammed, M. A., Abdulkareem, K. H., Mostafa, S. A., Ghani, M. K. A., Maashi, M. S., Garcia-Zapirain, B., ... & AL-Dhief, F. T. (2020). Voice Pathology Detection and Classification Using Convolutional Neural Network Model. *Applied Sciences*, *10*(11), 3723.

[13]   Zhang, H. (2004)"The optimality of Naive Bayes," *Proc. Seventeenth Int. Florida Artif. Intell. Res. Soc. Conf. FLAIRS 2004*, vol. 2, pp. 562–567.

[14]   Kharya, S., & Soni, S. (2016). Weighted naive bayes classifier: A predictive model for breast cancer detection. *International Journal of Computer Applications*, *133*(9), 32-37.

[15]   Xu, S., Li, Y., & Wang, Z. (2017). Bayesian multinomial Naïve Bayes classifier to text classification. In *Advanced multimedia and ubiquitous engineering* (pp. 347-352). Springer, Singapore.

[16]   Raschka, S. (2014). Naive bayes and text classification i-introduction and theory. *arXiv preprint arXiv:1410.5329*.

[17]   Wirth, R., & Hipp, J. (2000, April). CRISP-DM: Towards a standard process model for data mining. In *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining* (pp. 29-39). London, UK: Springer-Verlag.