Volume 9, Issue 06, June 2024 ISSN (online): 2456-0006
International Journal of Science Technology Management and Research
Available online at: www.ijstmr.com

Analysis and Impletation of Naive Bayes Classifier

Vaishnavi Patel
ME(Software Engineering) 4th semester
Computer Science and Engineering Department
Jawaharlal Institute of Technology Borawan (Khargone)

Mr. Kamlesh Patidar
Assistant Professor
Computer Science and Engineering Department
Jawaharlal Institute of Technology
Borawan (Khargone) M.P.

Abstract: ***The Naïve Bayes classifier is a supervised machine learning algorithm. which is used for classification tasks, such as text classification. Naive Bayes is a classification technique based on Bayes' Theorem with an assumption that all the features that predict the target value are independent of each other. It calculates each class's probability and then picks the one with the highest probability. Naive Bayes With "naive" assumption of independence among predictors It works with huge data and is mostly used to solve text kinds of data. Examples: Email classification Twitter sentiment analysis, etc. Bayes theorem is an indispensable law of probability, allowing you to deductively quantify unknown probabilities Bayes' Theorem allows you to update the predicted probabilities of an event by incorporating new Information Bayes' Theorem was named after 18th-century mathematician Thomas Bayes It often is employed in finance in calculating or updating risk evaluation The theorem has become a useful element in the implementation of machine learning. In this paper we implement the Naïve Bayes classifier for real life dataset and analysis its accuracy***

Keywords: Classifier, Machine learning , Probability, Accuracy, Prediction

I. INTRODUCTION

The Naive Bayes classifier separates data into different classes according to the Bayes' Theorem, along with the assumption that all the predictors are independent of one another. It assumes that a particular feature in a class is not related to the presence of other features.
For example, you can consider a fruit to be a watermelon if it is green, round and has a 10-inch diameter. These features could depend on each other for their existence, but each one of them independently contributes to the probability that the fruit under consideration is a watermelon. That's why this classifier has the term 'Naive' in its name.
The Naïve Bayes algorithm is composed of two words "Naïve" and "Bayes". It is named like this because it presumes the occurrence of a certain feature that is independent of the other features' occurrence. For example, if the fruit is recognized based on shape, colour, and taste, then the sweet fruit is recognized as an apple. So, every feature independently contributes to identifying that it is an apple without relying on each other. This algorithm is quite popular because it can even outperform highly advanced classification techniques. Moreover, it's quite simple, and you can build it quickly. Here's the Bayes theorem, which is the basis for this algorithm:

$$P(c \mid x) = P(x \mid c)\, P(c)/P(x)$$

In this equation, 'c' stands for class, and 'x' stands for attributes. P(c/x) stands for the posterior probability of class according to the predictor. P(x) is the prior probability of the predictor, and P(c) is the prior probability of the class. P(x/c) shows the probability of the predictor according to the class.

II. PRINCIPLE OF NAIVE BAYES CLASSIFIER

A Naive Bayes classifier is a probabilistic machine learning model that's used for classification task. The crux of the classifier is based on the Bayes theorem.
Bayes Theorem:

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

Using Bayes theorem, we can find the probability of **A** happening, given that **B** has occurred. Here, **B** is the evidence and **A** is the hypothesis. The assumption made here is that the predictors/features are independent. That is presence of one particular feature does not affect the other. Hence it is called naive.
Example:
Let us take an example to get some better intuition. Consider the problem of playing golf. The dataset is represented as below.

We classify whether the day is suitable for playing golf, given the features of the day. The columns represent these features and the rows represent individual entries. If we take the first row of the dataset, we can observe that is not suitable for playing golf if the outlook is rainy, temperature is hot, humidity is high and it is not windy. We make two assumptions here, one as stated above we consider that these predictors are independent. That is, if the temperature is hot, it does not necessarily mean that the humidity is high. Another assumption made here is that all the predictors have an equal effect on the outcome. That is, the day being windy does not have more importance in deciding to play golf or not.

	OUTLOOK	TEMPERATURE	HUMIDITY	WINDY	PLAY GOLF
0	Rainy	Hot	High	False	No
1	Rainy	Hot	High	True	No
2	Overcast	Hot	High	False	Yes
3	Sunny	Mild	High	False	Yes
4	Sunny	Cool	Normal	False	Yes
5	Sunny	Cool	Normal	True	No
6	Overcast	Cool	Normal	True	Yes
7	Rainy	Mild	High	False	No
8	Rainy	Cool	Normal	False	Yes
9	Sunny	Mild	Normal	False	Yes
10	Rainy	Mild	Normal	True	Yes
11	Overcast	Mild	High	True	Yes
12	Overcast	Hot	Normal	False	Yes
13	Sunny	Mild	High	True	No

According to this example, Bayes theorem can be rewritten as:

$$P(y|X) = \frac{P(X|y)P(y)}{P(X)}$$

The variable **y** is the class variable (play golf), which represents if it is suitable to play golf or not given the conditions. Variable **X** represents the parameters/features.
X is given as,

$$X = (x_1, x_2, x_3, ..., x_n)$$

Here x_1,x_2….x_n represent the features, i.e they can be mapped to outlook, temperature, humidity and windy. By substituting for **X** and expanding using the chain rule we get,

$$P(y|x_1, ..., x_n) = \frac{P(x_1|y)P(x_2|y)...P(x_n|y)P(y)}{P(x_1)P(x_2)...P(x_n)}$$

Now, we can obtain the values for each by looking at the dataset and substitute them into the equation. For all entries in the dataset, the denominator does not change, it remain static. Therefore, the denominator can be removed and a proportionality can be introduced.

$$P(y|x_1, ..., x_n) \propto P(y) \prod_{i=1}^{n} P(x_i|y)$$

In our case, the class variable(**y**) has only two outcomes, yes or no. There could be cases where the classification could be multivariate. Therefore, we need to find the class **y** with maximum probability.

$$y = argmax_y P(y) \prod_{i=1}^{n} P(x_i|y)$$

Using the above function, we can obtain the class, given the predictors.

III. ADVANTAGES AND LIMITATIONS OF NAIVE BAYES

The Naive Bayes is a popular algorithm due to its following advantages:

- This algorithm works very fast and can easily predict the class of a test dataset.
- You can use it to solve multi-class prediction problems as it's quite useful with them.
- Naive Bayes classifier performs better than other models with less training data if the assumption of independence of features holds.

- If you have categorical input variables, the Naive Bayes algorithm performs exceptionally well in comparison to numerical variables.
- It can be used for Binary and Multi-class Classifications.
- It effectively works in Multi-class predictions.

If your test data set has a categorical variable of a category that wasn't present in the training data set, the Naive Bayes model will assign it zero probability and won't be able to make any predictions in this regard. This phenomenon is called 'Zero Frequency,' and you'll have to use a smoothing technique to solve this problem.

- This algorithm is also notorious as a lousy estimator. So, you shouldn't take the probability outputs of 'predict_proba' too seriously.
- It assumes that all the features are independent. While it might sound great in theory, in real life, you'll hardly find a set of independent features.

IV. PROBLEM STATEMENT

Multiple Linear Regression analysis is the study of more than two variables in an attempt to find a relationship, or correlation. A regression line is a straight line that attempts to predict the relationship between two points, also known as a trend line or line of best fit. Multiple Linear Regression is a prediction when a variable (y) is dependent on more than one independent variable (x) based on the regression equation of a given set of data.

In the proposed work we have taken data of advertisement and sales. Advertisement on TV ,Radio and Newspaper and check the effect on sales.. Based on sales and advertiser various media (TV, Radio and Newspaper) we found that sales has been affected during advertisement on TV as compared to other two media To predict future stock index price by these two factors we used Multiple Linear Regression. We have the following problems.

1. How to select following variables
   - Dependent Variable
   - Independent Variable(s)
   - Intercept
   - Coefficients
2. How the model Select **data from the past** to learn what's the relationship.
3. How minimize the predicted error for regression analysis.
4. We need to check that a linear relationship exists between the dependent variable and the independent variable/s
5. We need to check that a linear relationship exists between the:
   Sales and advertisement on TV both have positive are negative relationship
   Sales and advertisement on Radio both have positive are negative relationship
   Sales and advertisement on Newspaper both have positive are negative relationship

V. LITERATURE SURVEY

In 2020 Krish Shah et al proposed "Real Time Diabetes Prediction using Naïve Bayes Classifier on Big Data of Healthcare" Diabetes is a chronic disease, with numerous cases enrolled annually. The number of deaths caused by diabetes has been expanding each year, and it is crucial to anticipate the factor so that they can be relieved at the soonest guaranteeing the patient's life is saved. This prediction is effectively acquired by utilizing Naïve Bayes Classifier. This algorithm classifies, based on the indications of whether an individual has diabetes or not. This model achieved an accuracy of around 81%. The proposed system also supports live streaming of data input, where the results are obtained in real-time for the entered patient they used two classifiers based on machine learning algorithm Naïve Bayes, Multinomial Naïve Bayes Classifier and Gaussian Naïve Bayes Classifier for experimentation to predict Diabetes disease. The two classifiers thus built have been compared based on their accuracy value. Another performance evaluation method was classifier accuracy measure which included TP-rate, FP-rate, precision, recall, F-Measure. The overall performance of Gaussian Naïve Bayes Classifier to predict diabetes disease is better than Multinomial Naïve Bayes Classifier. Hence the viability of the proposed model is clearly portrayed throughout the experimental results[1]

In 2021 Hong Chen et al proposed "Improved naive Bayes classification algorithm for traffic risk management" Naive Bayesian classification algorithm is widely used in big data analysis and other fields because of its simple and fast algorithm structure. Aiming at the shortcomings of the naive Bayes classification algorithm, they used feature weighting and Laplace calibration to improve it, and obtains the improved naive Bayes classification algorithm. Through empirical research, it is found that the improved naive Bayes classification algorithm can greatly improve the correct rate of discrimination analysis from 49.5 to 92%. Through robustness analysis, the improved naive Bayes classification algorithm has higher accuracy they improves the algorithm by using the feature weighting and Laplace calibration and obtains the improved naive Bayesian classification algorithm. The results show that when the sample size is large, the improved naive Bayesian classification algorithm has a high accuracy of 99% and is very stable. When the sample attribute is less than 400, the accuracy rate is over 95%, and when the sample attribute is greater than 600, the accuracy rate of discrimination decreases to about 50%, and the trend is stable; when the number of categories is less than 24, the accuracy rate of discrimination analysis is maintained at least 95%, and the trend is stable; when the number is more than 60, the accuracy of discrimination is reduced to zero rapidly[2].

In 2021 Sushma S A et al proposed "Comparative Study of Naive Bayes, Gaussian Naive Bayes Classifier and Decision Tree Algorithms for Prediction of Heart Diseases". Nowadays death due to heart disease has been common in the world. It has become a hard task for the medical practitioners to diagnose in the initial stage and requires more expertise and demand in the medical field for prediction. Designing an automated system by using machine learning algorithm will improve the medical efficiency and also reduce the cost. They

designed an automated system that can be used for efficiently predicting the results which give information about the risks need to be faced by the patients with respect to heart diseases by using the parameter available in the dataset. They extracted the hidden patterns from the parameters by applying data mining techniques. Since the heart data is too massive and complex for analysis using traditional techniques, they used machine learning algorithm for computation using the parameters available in the dataset and produce accurate prediction of heart disease. Machine Learning Prediction techniques like Naive Bayes Classifier, Gaussian Naïve Bayes Classifier and Decision tree can be used to analyze and predict heart diseases[3].

In 2021 Jasna P. S. et al proposed "D-Predict: Disease Prediction Using Gaussian Naive Bayes Algorithm". The application of machine learning in the field of medical diagnosis is increasing gradually. This can be contributed primarily to the improvement in the classification and recognition systems used in disease diagnosis which is able to provide data that aids medical experts in early detection of fatal diseases and therefore, increase the survival rate of patients significantly. The Existing systems for disease prediction are narrowed down one or a few diseases or medical conditions that are very crucial in the healthcare field. So, the objective of the D-predict is that it will predict the diseases that are seen in old aged people using the concept of natural language processing and machine learning. According to the D-predict when user give their symptoms as input in the form of speech, after NLP processing the key features are extracted and using Naive Bayes classifier the disease is predicted. The system can diagnose and predict disease efficiently at an earlier stage with reasonable accuracy D-predict is very much useful in old aged citizen's day to day life and it is mainly more important for the healthcare sector, because they are the one that daily uses these systems to predict the disease on the basis of the symptoms. The project is designed in such a way that the system takes symptoms from the user as voice and produces output i.e, predict disease. Disease Prediction is done using NLP and ML algorithm. The User can give their symptom as input in the form of speech. After NLP processing the key features are extracted and finally the disease prediction is done using Gaussian Naive bayes classifier [4].

In 2022 M. Vijay Anand et al proposed "Gaussian Naive Bayes Algorithm: A Reliable Technique Involved in the Assortment of the Segregation in Cancer". Cancer is a disease caused by uncontrollable cell growth. Disease is a constant subject of concern due to unavailability of treatment at a severe level. Patients who have suffered from the disease have the chance of getting saved if this fatal illness is identified in the beginning stage. Survival chance will be very low if it is detected in the final stage of cancer. As the patients could not survive in their last stage, to cure their disease, an early diagnosis is a key issue and is vital. For the classification of cancer, Gaussian Naive Bayes is implemented in this work. By exerting it on two datasets, the algorithm is tested, in which the Wisconsin Breast Cancer Dataset (WBCD) is considered as earliest one and the next one is the Lung Cancer Dataset. 'e assessment result of the suggested algorithm attained 90% accuracy in the prediction of lung cancer, and in predicting breast cancer, the accuracy is 98%. To identify the disorganized value in the classification for normalization, the proposed work presents an effective approach, which deserves to be zero. For the prediction of cancer based upon Gaussian distribution, the algorithm of Naive Bayes is used. Among the two types of cancer named lung cancer as well as breast cancer, the model was implemented, and to list the patient's details, an interface has been intended, which predicts the class of cancer. Therefore, high classification performance has been achieved by the proposed work[5].

In 2022 Chingmuankim et al proposed "A comparative study of Naive Bayes Classifiers with improved technique on Text Classification". Naive Bayes classifier is a linear classifier based on Bayesian theorem that make use of conditional probabilities where the class of the comments in the dataset are decided by the result obtained from the product of prior probability and likelihood of the class. These textual features are represented into a vector form using different feature extraction techniques of Natural Language Processing such as Bag of Words and Term Frequency Inverse Document Frequency for the purpose of computation. Naive Bayes is also known as an eager learner due to its quick capability to learn. It has been widely used for classification of text. One of the well-known limitations of this classifier is the independent assumption of features which creates a possibility for the occurrence of redundant features. Moreover, the significance of semantics is completely ignored. Despite this limitation, this classifier has performed very well. Earlier research has been conducted on comparison between two models of Naive Bayes, namely Bernoulli and Multinomial Naive Bayes, distribution of the datasets, errors such as system level error and weight error has been analyzed. The objective of this work is to test imbalance dataset on the three types of Naive Bayes Model, compare the performance result and see the impact it has on the model by applying our improved technique that combines TF-IDF with n-gram language modelling. They discussed in detail the experimental results and the methodology used for carrying out this research. Our experimental result shows a 2-3% increase in accuracy compared than the old model. TF-IDF is more preferable than BOW due to its inability to understand semantics of the textual data, which is obsolete in the real word[6].

In 2023 Daniel Jurafsky et al proposed "Naive Bayes and Sentiment Classification" Many language processing tasks involve classification, although luckily our classes are much easier to define than those of Borges. They introduced the naive Bayes algorithm and apply it to text categorization, the task of assigning a label or text categorization category to an entire text or document. They focus on one common text categorization task, sentiment analysis, the ex- sentiment analysis traction of sentiment, the positive or negative orientation that a writer expresses toward some object. A review of a movie, book, or product on the web expresses the author's sentiment toward the product, while an editorial or political text expresses sentiment toward a candidate or political action. Extracting consumer or public sentiment is thus relevant for fields from marketing to politics. The simplest version of sentiment analysis is a binary classification task, and the words of the review provide excellent cues. Consider, for example, the following phrases extracted from positive and negative reviews of movies and restaurants. Words like great, richly, awesome, and pathetic, and awful and ridiculously are very informative cues[7].

In 2023 Kalakonda Shashank et al proposed " In our day-to-day life reviews plays a crucial role in every business aspect. To increase business scale, the product must have positive reviews from the users. There are hundreds of similar products in every business. To avoid his/her difficulty customers always check the reviews whether it satisfies them or not, and they buy the products only after it satisfies them. So, owners always should observe their reviews and do changes according to them. If there are below a hundred reviews, the owner

had possibility to read those reviews and determine whether they are positive or not. If there are thousands of reviews owner cannot read all of them. So, to overcome this problem, we are using Machine Learning concept Gaussian naïve bayes and predict the positive or negative reviews. A dataset consisting of reviews taken as inputs and preprocessing the data and applying the algorithm. A spreadsheet consisting of reviews with their respective values (0/1 1 for positive and 0 for negative) is the output. They observed the results of this paper that has done on reviews taken from restaurant by using Gaussian Naive bayes and found can able to understand the sentiment from customers. The accuracy given this model is 72.77%. Further research can be done by taking more number of reviews or variety of data or by using different models to increase the accuracy[8].

In 2023 M. Vedaraj et al proposed "Early Prediction of Lung Cancer Using Gaussian Naive Bayes Classification Algorithm". The early prediction of lung cancer is of utmost importance for improving patient survival rates. They proposed the implementation of the Gaussian Naive Bayes (GNB) classification algorithm to detect lung cancer at its nascent stages. The researchers assess the performance of the GNB algorithm by employing a lung cancer dataset obtained from the University of California, Irvine (UCI). To gauge the effectiveness of GNB, its results are compared against other popular ML techniques such as K-Nearest Neighbors (KNN), Support Vector Machines (SVM), and the J48 algorithm (a variant of the C4.5 decision tree algorithm). Notably, the performance analysis reveals that the GNB algorithm achieves an impressive 98% accuracy in predicting lung cancer. This signifies the promising potential of GNB for accurate and early-stage detection of lung cancer. By leveraging the distinctive characteristics of the Gaussian Naive Bayes algorithm and utilizing the lung cancer dataset, the researchers successfully demonstrated its efficacy in achieving a high level of accuracy. This research contributes to the on-going efforts in improving lung cancer diagnosis and emphasizes the significance of early prediction in enhancing patient outcomes The proposed method for early prediction of lung cancer using the Gaussian Naive Bayes classification algorithm demonstrates promising results with high accuracy, precision, specificity, F1-score, and sensitivity[9].

VI. PRAPOSED APPROACH

Working of Naïve Bayes' Classifier.
A dataset consisting of weather conditions and the relevant target variable "GamePlay". We have to determine whether we must play or not according to the weather conditions of the particular day and based on the dataset. Follow these steps to solve this problem.

- Transform the given dataset into frequency tables.
- Generate a Likelihood table by determining the probabilities of the given features.
- Apply Bayes theorem to determine the posterior probability.

Now suppose your problem is: Should the player play if the weather is sunny.

VII. COMPARATIVE ANALYSIS

Comparing with confusion matrix for TP and FP.

Simple example of Confusion Matrix

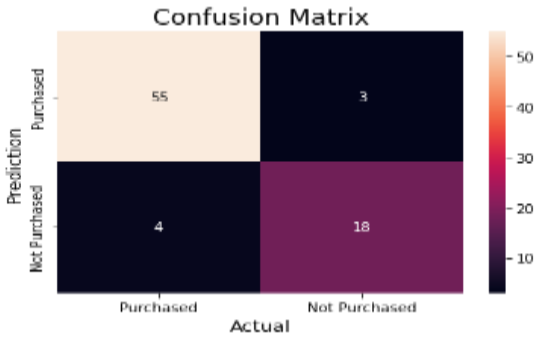

Table 2 Confusion matrix True Positives (TP) and False Positives (FP)

TP	FP
55	3

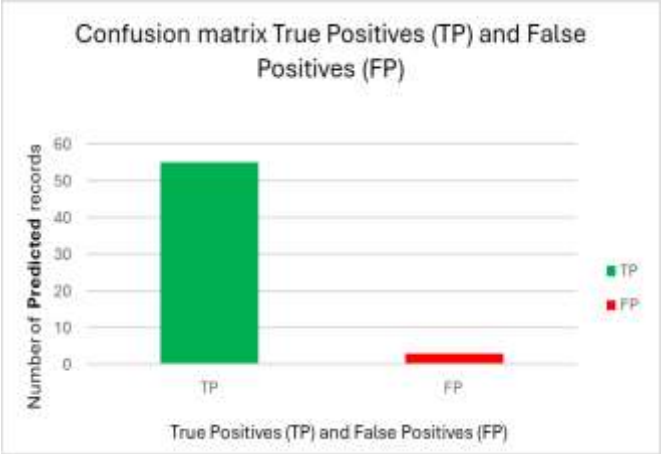

Figure 1 Confusion matrix True Positives (TP) and False Positives (FP)

VIII. CONCLUSION

Bayes theorem is given by an English statistician, philosopher, and Presbyterian minister named Mr. Thomas Bayes in 17th century. Bayes theorem provides thoughts in decision theory which is extensively used in important mathematics concepts such as Probability. Bayes theorem is widely used in Machine Learning where we need to predict classes precisely and accurately. An important concept of Bayes theorem named Bayesian method is used to calculate conditional probability in Machine Learning application that includes classification tasks. Simplified version of Bayes theorem (Naïve Bayes classification) is also used to reduce computation time and average cost of the projects. Naïve Bayes are robust to isolated noise points because such points are averaged out when estimating conditional probabilities from data

REFERENCE

1. Krish Shah, Rajiv Punjabi, Priyanshi Shah, Dr Madhuri Rao Real Time Diabetes Prediction using Naïve Bayes Classifier on Big Data of Healthcare International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 07 Issue: 05 | May 2020 www.irjet.net p-ISSN: 2395-0072
2. Hong Chen , Songhua Hu , Rui Hua and Xiuju Zhao Hong Chen , Songhua Hu , Rui Hua and Xiuju Zhao Chen et al. EURASIP Journal on Advances in Signal Processing (2021) 2021:30 https://doi.org/10.1186/s13634-021-00742-6
3. 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 (IJRASET) ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429 Volume 9 Issue III Mar 2021- Available at www.ijraset.com
4. 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 (IJCESR)
5. 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, Article ID 2436946, 7 pages https://doi.org/10.1155/2022/2436946.
6. Naulak, Chingmuankim A comparative study of Naive Bayes Classifiers with improved technique on Text Classification https://www.techrxiv.org/articles/preprint/ 10.36227/techrxiv.19918360.v1.
7. 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.
8. Kalakonda Shashank1, Anumandla Sahithya2, Shaik Shakeel3, Dr. R. Lakshmipriya4 Reviews analysis Using gaussian naïve bayesin machine Learning International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 10 Issue: 01 | Jan 2023 www.irjet.net p-ISSN: 2395-0072
9. M. Vedaraj, 2C.S. Anita, 3A. Muralidhar, 4V. Lavanya, 5K. Balasaranya, 6P. Jagadeesan Early Prediction of Lung Cancer Using Gaussian Naive Bayes Classification Algorithm International Journal of INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING ISSN:2147-67992147-6799 www.ijisae.org Original Research Paper International Journal of Intelligent Systems and Applications in Engineering.
10. 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)
11. Nafizatus Salmi and Zuherman Rustam Naïve Bayes Classifier Models for Predicting the Colon Cancer 9th Annual Basic Science International Conference 2019 (BaSIC 2019) IOP Conf. Series: Materials Science and Engineering 546 (2019) 052068 IOP Publishing doi:10.1088/1757-899X/546/5/052068
12. 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)
13. 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
14. Amitha, Merin Meleet A System for Recommendation of Medication Using Gaussian Naïve Bayes Classifier International Journal of Innovative Research in Computer Science & Technology (IJIRCST) ISSN: 2347-5552, Volume-7, Issue-3, May-2019 DOI: 10.21276/ijircst.2019.7.3.13
15. 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 © Universiti Tun Hussein Onn Malaysia Publisher's Office JSCDM Journal homepage: http://penerbit.uthm.edu.my/ojs/index.php/jscdm Journal of Soft Computing and Data Mining e-ISSN : 2716-621X