BAREKENG: Journal of Mathematics and Its Applications
September 2023 Volume 17 Issue 3 Page 1595–1604
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol17iss3pp1595-1604

DETERMINATION OF BANK INDONESIA SCHOLARSHIP RECIPIENTS USING NAÏVE BAYES CLASSIFIER

Fera Malianis Febri[1], Devni Prima Sari[2*]

[1,2]*Mathematics Departement, Faculty of Mathematics and Natural Sciences, Padang State University*
Prof. Dr. Hamka Street, Air Tawar, Kota Padang, 25131, West Sumatera

*Corresponding author's e-mail: * **devniprimasari@fmipa.unp.ac.id***

ABSTRACT

A scholarship is a grant given to students as financial aid for education. One of the most sought-after scholarships is the scholarship from Bank Indonesia. Currently, the selection process for Bank Indonesia scholarship recipients still involves verifying the completeness of the administrative documents of the prospective recipients. Manual administrative verification requires a long time for data processing and re-verification. Therefore, there is a need for a data classification system to assist in the decision-making process for Bank Indonesia scholarship recipients. This study aims to implement the naïve Bayes classifier method to classify Bank Indonesia scholarships accurately. The variables used include gender, semester, parental income, grade point average (GPA), achievement, organizational activity, and the number of dependents. This research found that the naïve Bayes classifier method for classifying Bank Indonesia scholarship recipients can be done accurately with an accuracy rate of 86,84%.

Article History:

Received: 12th April 2023
Revised: 8th August 2023
Accepted: 16th August 2023

Keywords:

Barekeng journal;
Mathematics topic;
Alphabetic

How to cite this article:

F. M. Febri and D. P. Sari, "DETERMINATION OF BANK INDONESIA SCHOLARSHIP RECIPIENTS USING NAÏVE BAYES CLASSIFIER," *BAREKENG: J. Math. & App.*, vol. 17, iss. 3, pp. 1595-1604, September, 2023.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

Research Article · Open Access

1. INTRODUCTION

Education is one of the supporting factors for creating quality human resources. Education is changing people's attitudes and behavior to mature humans through teaching, training, and evaluation [1]. Education is expected to change and develop one's mindset and intelligence. To optimize the utilization and processing of existing resources [2], [3]. However, the implementation of quality education requires a large amount of money. Therefore, every student who excels and whose parents are underprivileged is entitled to a scholarship.

These scholarships can come from government, private or foundation institutions [4]. Scholarship providers generally provide some criteria that must be met, for example, parents' income, number of dependents, GPA, achievement of scholarship recipients, living costs, and others [3]. One of the scholarships offered is a scholarship from Bank Indonesia. Bank Indonesia Scholarships are scholarships issued by Bank Indonesia as a form of Bank Indonesia's dedication to the world of education. Bank Indonesia Scholarships are intended for economically disadvantaged students who have at least completed 40 credits or are currently in their 3rd semester. The student also a maximum of 23 years of age when designated as scholarship recipients, are not now receiving scholarships from other institutions, and are willing to play an active role in managing and developing the Indonesian Generasi Baru Indonesia (GenBI)[5].

The selection procedure for a Bank Indonesia scholarship has several standards that students must fulfill. The selection process for scholarship recipients Bank Indonesia still selects the completeness of administrative documents for scholarship recipients. Manually selecting administrative completeness necessitates a lengthy data processing and re-verification process. Consequently, a data classification system is required to aid in selecting Bank Indonesia scholarship recipients [6]. This research differs from earlier research is Visual Basic and Weka was the program utilized by earlier studies while conducting this study, RStudio software was used.

A classification approach based on probabilities is the naive Bayes classifier method. The naive Bayes classifier method's basic tenet is to predict opportunities for the future based on historical data [7], [8], [9]. The naive Bayes classifier approach, also known as conditional independence, assumes that a particular class's impact is independent of other variables [1]. The researcher wants to talk about how the naive Bayes classifier approach was used to determine who would receive a Bank Indonesia scholarship in light of the aforementioned background.

2. RESEARCH METHODS

This study is an example of applied research, which aims to address present or upcoming issues. The secondary data utilized was obtained from Bank Indonesia. The naive Bayes classifier approach was employed in this study.

2.1 Research Variable

The data is from West Sumatra students who registered for the Bank Indonesia scholarship in 2022 [6]. The data was obtained directly from the West Sumatra branch of Bank Indonesia. The data used were 570 students and eight variables. The variables consist of 7 independent variables and one dependent variable. Table 1 shows the specifics of each variable.

Table 1. Research Variables

Variable Type	Variable Name	Description	Category
Independent	X_1	Gender	1: Male | 2: Female
	X_2	Semester	1: Semester 3 | 2: Semester 5 3: Semester 7 | 4: Semester 9
	X_3	Parental Income	1: Less than IDR 1,000,000 | 2: IDR 1,000,000 - Rp 3,000,000 | 3: More than IDR 3,000,000
	X_4	GPA	1: Less than 3.00 | 2: 3.00 - 3.25 3: 3.26 - 3.50 | 4: More than 3.50
	X_5	Achievements	1: No achievement | 2: Local level 3: Provincial level | 4: National level | 5: International level

Variable Type	Variable Name	Description	Category
	X_6	Organizational Involvement	1: Not involved | 2: Involved
	X_7	Number of Dependents	1: Less than four people | 2: 4 - 6 people | 3: More than six people
Dependent	Y	Final Decision	1: Accepted | 2: Rejected

2.2 Data Mining

Data mining uses one or more machine learning algorithms to examine and extract knowledge automatically. One of the crucial steps in Knowledge Discovery Databases (KDD) is data mining. The KDD method entails several stages to convert unusable data into useful information. This is how KDD works [10], [11]:

a. Data Cleaning
b. Data Integration
c. Data Selection
d. Data Transformation
e. Mining Process
f. Pattern Evaluation
g. Knowledge Presentation

2.3 Naïve Bayes Classifier

The Naive Bayes classifier (NBC) is a statistical classifier that can forecast the likelihood that a class will include members. Thomas Bayes, a British physicist, developed the Bayes theory based on the naive Bayes classifier. Based on historical data, this theorem can forecast potential future outcomes [12]. Based on probability and graph theory, the Bayesian Network is a paradigm for modeling joint distributions in graphs. A Directed Acyclic Graph (DAG) is the fundamental building block of the Bayesian Network construction. The Conditional Probability Table (CPT), the quantitative portion of the collection of conditional probability distributions for each variable based on the graph [13], [14], [15] is the following element. The Bayes theorem has the following generic form:

$$P(Y|X) = \frac{P(Y)P(X|Y)}{P(X)} \tag{1}$$

Where:

$P(Y|X)$: the probability of hypothesis Y given condition X

$P(X|Y)$: the possibility of X given condition Y

$P(Y)$: the case of hypothesis Y

$P(X)$: the probability of X

In the Naive Bayes classifier method, all attributes are assumed to be equally essential or have equal priority. Also, the attribute values in a class are independent of the values in other features [16], [17], [18]. Thus, Bayes' theorem in Equation (1) can be adjusted as follows:

$$P(Y|X_1, X_2, \dots, X_n) = \frac{P(Y)P(X_1, X_2, \dots, X_n|Y)}{P(X_1, X_2, \dots, X_n)} \tag{2}$$

Where:

$X_1 \dots X_n$: Variables corresponding to classifying traits

Y : Variable denoting class

$P(Y)$: Probability of the occurrence of class Y (Prior Probability)

$P(X_1 \dots X_n)$: Probability of the occurrence of characteristics $X_1 \dots X_n$ (Evidence)

$P(Y \mid X_1 \dots X_n)$: Probability of the class Y occurring based on $X_1 \dots X_n$ attributes (Posterior Probability)

$P(X_1 \dots X_n \mid Y)$: Probability of characteristics $X_1 \dots X_n$ based on the occurrence of class Y (Likelihood)

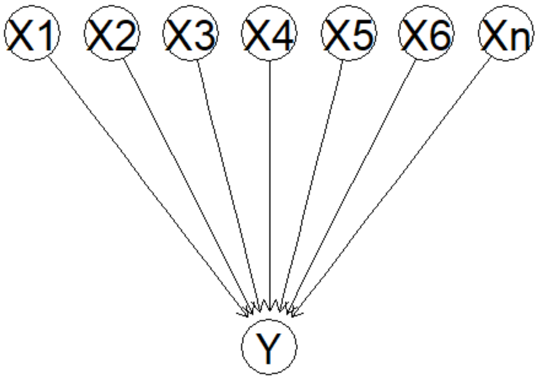

Figure 1. **DAG Naïve Bayes**

Depending on **Figure 1** the chance of occurrence of the random variables $X_1, \dots, X_n, Y$ is known, hence the following definition of conditional probability is used to determine the conditional probability of occurrence of Y against X.

$$P(X_1 \cap \dots \cap X_n \cap Y) = P(Y|X_1 \cap \dots \cap X_n)P(X_1) \dots P(X_n) \tag{3}$$

Or

$$P(X_1 \cap \dots \cap X_n \cap Y) = P(X_1 \cap \dots \cap X_n|Y)P(Y) \tag{4}$$

From the equations following formula can be obtained.

$$P(Y|X_1 \cap \dots \cap X_n) = \frac{P(X_1 \cap X_2 \cap \dots \cap X_n|Y)P(Y)}{P(X_1)P(X_2) \dots P(X_n)} \tag{5}$$

Since the random variables $X_1, X_2, \dots, X_n$ are independent events, Dawid [19] states the following.

$$P(X_1 \cap X_2 \cap \dots \cap X_n|Y) = P(X_1|Y)P(X_2|Y) \dots P(X_n|Y) \tag{6}$$

The value of evidence for a single sample is always constant for each class. The posterior value will then be compared with the posterior values of other classes to determine which class a sample will be classified into. The Bayes formula is further explained by expanding $(Y, X_1, \dots, X_n)$ using the multiplication rule as follows:

$$P(Y|X_1 \dots X_n) = P(Y).P(X_1|Y).P(X_2|Y, X_1) \dots P(X_n|Y, X_1, X_2, \dots, X_{n-1}) \tag{7}$$

Based on the equation above, this strong independence assumption simplifies the probability condition and makes the calculation possible. **Equation (7)** can also be denoted as follows [20], [21]:

$$P(Y|X_1, X_2, \dots, X_n) \propto P(Y)P(X_1|Y)P(X_2|Y) \dots P(X_n|Y)$$
$$\propto P(Y) \prod_{i=1}^{n} P(X_i|Y) \tag{8}$$

Where $\propto$ denotes proportionality

This means that under the above independence assumptions, the conditional distribution over the class variable Y is [22]:

$$P(Y|X) = \frac{P(Y) \prod_{i=1}^{n} P(X_i|Y)}{P(X)} \tag{9}$$

Where the evidence

$$P(X) = \sum_k P(Y_k)P(X|Y_k) \tag{10}$$

In Naïve Bayes Classifier we need to maximize the probability value of each class, which is expressed as the HMAP:

$$H_{MAP} = argmax\ P(Y|X) = argmax\ P(Y)\prod_{i=1}^{n} P(X_i|Y) \tag{11}$$

2.4 Classifier Accuracy Test

The performance measurement of a classification system is carried out to determine the level of accuracy of the system that has been created. The performance measurement of the classification can be carried out using a confusion matrix **[23]**, **[24]**. A confusion matrix is a table that records the results of the classification work. The calculation of the confusion matrix is shown in the following **Table 2**.

Table 2. Confusion Matrix

Forecast Value		True Value	
		True	False
	True	True Positive	False Positive
	False	False Negative	True Negative

Explanation:

TN: Data with an actual negative value and predicted negative value

TP: Data with a substantially positive value and predicted positive value

FN: Data with a significant positive value and predicted negative value

FP: Data with a significant negative value and predicted positive value

Classification performance can be evaluated by considering the following measures:

a. Accuracy is a measure of the classification prediction accuracy on the dataset. It can be calculated by knowing the data is classified correctly. To calculate accuracy, the following formula is used [10]:

$$Accuracy = \frac{TN + TP}{TN + TP + FN + FP} \tag{12}$$

b. Error rate is the proportion of classification that makes prediction errors, calculated using the following equation **[10]**:

$$Error\ rate = \frac{FN + FP}{TN + TP + FN + FP} \tag{13}$$

2.5 Research Stages

The following steps will be taken to solve the problem:

a. Identify the problem to be addressed in this research: determining the recipients of Bank Indonesia scholarships using the naïve Bayes classifier method.

b. Review and study books, journals, and other sources related to the naïve Bayes classifier method and the criteria for scholarship recipients at Bank Indonesia.

c. Retrieve data from Bank Indonesia.

d. Build the structure of the naïve Bayes classifier.

   i. Divide the data into two parts: testing and training data.

   ii. Based on the training data, calculate the testing data's prior probability $P(Y)$.

   iii. Calculate the probability of each variable (likelihood) for each class $(P(X_i|Y))$ on the testing data based on the training data.

iv. Calculate the multiplication of probabilities with the probability of attributes in each class $P(Y)P(X_i|Y)$

v. Find the maximum value of $P(Y|X) = \frac{P(Y)\prod_{i=1}^{n} P(X_i|Y)}{P(X)}$ in both classes.

vi. The highest value from the calculation is the prediction result.

e. Perform accuracy testing of the naïve Bayes classifier model using a confusion matrix.

f. If the accuracy is insufficient, repeat the steps to build the naïve Bayes classifier structure. If it is accurate, conclude.

3. RESULTS AND DISCUSSION

3.1 Split Data

Training and testing set of data are separated in classification. In machine learning, training data are used to develop a model or algorithm [25]. In the meantime, a model developed using training data is assessed using testing data for correctness. This testing is intended to assess the model's capacity to forecast target values from unobserved data [25]. The amount of data needed for training or testing has no set guidelines. In this study the propotion data used 80% for training and 20% for testing data. The breakdown of training and testing data is displayed in the table below.

Table 3. Training and Testing Data Proportion

Data	Amount
Training	456
Testing	114

3.2 Classification with Naive Bayes Classifier

The naive Bayes classifier method divides the classification process into three stages: prior probabilities, the likelihood of each independent variable in each group, and the multiplication of prior probabilities and the likelihood of each independent variable in each group.

a. Calculating Prior Probabilities

The training data are used to determine the prior probabilities in each group in the naive Bayes classifier method's classification procedure. In the 456 available training data, 360 scholarship applicants are accepted, so the prior probability of the first group is as follows:

$$P(Y = 1) = \frac{360}{456} = 0,789474$$

Thus, the prior probability of scholarship applicants being accepted is 0.789474. While 96 scholarship applicants are rejected, the prior probability of the second group is as follows:

$$P(Y = 2) = \frac{96}{456} = 0,210526$$

Thus, the prior probability of scholarship applicants being rejected is 0.210526. Table 4 displays the probabilities of applicants for scholarships.

Table 4. Prior Probabilities

	$P(Y)$
1	0,789473
2	0,210526

b. Calculating the Probability of Each Independent Variable

Next, the probability of each variable (likelihood) that influences the acceptance of Bank Indonesia scholarships will be calculated.

The probability values of the gender variable for each group are as follows:

$$P(X_1 = 1|Y = 1) = \frac{90}{360} = 0{,}250000$$

The probability of gender variable for male (1) declared as reject can be calculated as follows:

$$P(X_1 = 1|Y = 2) = \frac{41}{96} = 0{,}427083$$

Therefore, the probability of the male gender variable (1) declared as rejected is 0,427083. Repeat the process for each variable. The details are shown in **Table 5**.

Table 5. $\boldsymbol{P(X_i|Y)}$

X_1	Y	
	1	2
1	0,250000	0,427083
2	0,750000	0,572917
X_2		
1	0,138889	0,312500
2	0,405556	0,302083
3	0,430556	0,354167
4	0,025000	0,031250
X_3		
1	0,044444	0,104167
2	0,580556	0,635417
3	0,375000	0,260417
X_4		
1	0,002747	0,100000
2	0,260989	0,630000
3	0,271978	0,230000
4	0,464285	0,040000
X_5		
1	0,594520	0,722772
2	0,301369	0,168316
3	0,087671	0,059405
4	0,010958	0,039603
5	0,005479	0,009900
X_6		
1	0,063889	0,208333
2	0,936111	0,791667
X_7		
1	0,433333	0,468750
2	0,525000	0,437500
3	0,041667	0,093750

c. Calculating Prior Probability and Probability of Each Independent Variable Multiplication

The next step is to multiply the prior probability and probability of the independent variables (likelihood) to determine the class in the testing data. The calculation of the prior probability and probability of independent variables multiplication using **Equation (9)** is as follows:

$$P(Y|X_1, X_2, X_3, X_4, X_5, X_6, X_7) = P(Y)\prod_{i=1}^{7} P(X_i|Y)$$
$$= P(Y)P(X_1|Y)P(X_2|Y)P(X_3|Y)P(X_4|Y)P(X_5|Y)P(X_6|Y)P(X_7|Y)$$

In the first testing data, it is known that the scholarship recipient candidate is male, in semester 5, parents' income is between IDR 1,000,000 - IDR 3,000,000, has a GPA of more than 3.50, has not achieved any achievements, actively participates in organizations, and has less than four dependents.

The results of the prior probability multiplication and the probability of the independent variables based on above are as follows:

$$P(Y = 1|X_1, X_2, X_3, X_4, X_5, X_6, X_7)$$
$$= P(Y = 1) \times P(X_1 = 1|Y = 1) \times P(X_2 = 2|Y = 1) \times P(X_3 = 2|Y = 1) \times P(X_4 = 4|Y = 1) \times P(X_5 = 1|Y = 1) \times P(X_6 = 2|Y = 1) \times P(X_7 = 1|Y = 1)$$
$$= 0{,}789473 \times 0{,}250000 \times 0{,}405556 \times 0{,}580556 \times 0{,}464285 \times 0{,}594520 \times 0{,}936111 \times 0{,}43333$$
$$= 0{,}005203$$

The probability of data testing being accepted is 0.005203. While the probability of being rejected is:

$$P(Y = 2|X_1, X_2, X_3, X_4, X_5, X_6, X_7)$$
$$= P(Y = 2) \times P(X_1 = 1|Y = 2) \times P(X_2 = 2|Y = 2) \times P(X_3 = 2|Y = 2) \times P(X_4 = 4|Y = 2) \times P(X_5 = 1|Y = 2) \times P(X_6 = 2|Y = 2) \times P(X_7 = 1|Y = 2)$$
$$= 0{,}21052 \times 0{,}427083 \times 0{,}302083 \times 0{,}635417 \times 0{,}040000 \times 0{,}722772 \times 0{,}791667 \times 0{,}468750$$
$$= 0{,}000185$$

After calculating the posterior probabilities for both groups, the maximum value for each group is determined to determine the classification result. Based on the calculation results, the group with the highest posterior probability is the group where the candidate is declared accepted, which is 0.005203.

3.3 Classification Accuracy Test

To evaluate the results of the classification on naïve Bayes method, the accuracy level is measured. Based on **Table 6** it can be calculated accuracy and error rate as follows:

$$Accuracy = \frac{86 + 13}{86 + 4 + 13 + 11} = \frac{99}{114} = 0{,}8684$$

and

$$Error\ Rate = \frac{11 + 4}{86 + 4 + 13 + 11} = \frac{15}{114} = 0{,}1315$$

The tests' accuracy was 86.84%, and their error rate was 13.15 %. **Table 6** of the categorization results shows the number of correctly and wrongly categorised objects.

Table 6. Confusion Matrix Result

Predict		Actual 1	Actual 2
	1	86	11
	2	4	13

4. CONCLUSIONS

The following inferences can be made considering the data analysis and discussion:

a. The classification of Bank Indonesia scholarship recipients using the naïve Bayes classifier method resulted in 86 correctly classified out of 90 declared eligible and four others. Meanwhile, out of 24 unqualified, 13 were correctly classified as not qualified, and 11 were incorrectly classified.

b. The measurement of the classification accuracy using the confusion matrix resulted in an accuracy rate of 86.84%. This indicates that the naïve Bayes classifier method provides good classification predictions.

REFERENCES

[1] A. Hadianto, B. Setya Rintyarna, and L. Ali Muharom, "Klasifikasi Mahasiswa Penerima Beasiswa Dengan Metode Naive Bayes," 2009.

[2] Asnimar, A. Achmad, Yuyun, A. Iskandar, and Mansyur, "Classification Of Prospective Scholarship Recipients Kartu Indonesia Pintar (KIP) With Decision Tree Algorithm And Naïve Bayes," *Inspir. J. Teknol. Inf. dan Komun.*, vol. 12, no. 2, pp. 118–129, 2022, doi: 10.35585/inspir.v12i2.6.

[3] H. Kurniawan, I. R. Munthe, and B. Bangun, "Analysis of Naive Bayes Methods on Scholarship Admissions at Smks Al-Wasliyah 2 Merbau," *J. Mantik*, vol. 6, no. 1, pp. 364–373, 2022.

[4] R. Hegde, V. Anusha G, S. Madival, S. Sowjanya H, and U. Sushma, "A Review on Data Mining and Machine Learning Methods for Student Scholarship Prediction," *Proc. - 5th Int. Conf. Comput. Methodol. Commun. ICCMC 2021*, no. ICCMC, pp. 923–927, 2021, doi: 10.1109/ICCMC51019.2021.9418376.

[5] Ismuato'illah, "Pengaruh Pemberian Beasiswa Bank Indonesia Terhadap Motivasi Berprestasi Mahasiswa Ditinjau Dari Perspektif Ekonomi Islam," *Skripsi*, 2020.

[6] M. I. A. Putera and M. G. L. Putra, "Rancang Bangun Sistem Pendukung Keputusan Penentuan Calon Penerima Beasiswa Menggunakan Metode Simple Additive Weighting Pada Kpw Bank Indonesia Balikpapan," *J. Ilm. Tek. Inf.*, vol. 14, no. 2, pp. 110–120, 2020, [Online]. Available: https://ejournal.unisbablitar.ac.id/index.php/antivirus/article/view/1213/896

[7] S. Chen, G. I. Webb, L. Liu, and X. Ma, "A novel selective naïve Bayes algorithm," *Knowledge-Based Syst.*, vol. 192, p. 105361, 2020, doi: 10.1016/j.knosys.2019.105361.

[8] S. Dey, S. Wasif, D. S. Tonmoy, S. Sultana, J. Sarkar, and M. Dey, "A Comparative Study of Support Vector Machine and Naive Bayes Classifier for Sentiment Analysis on Amazon Product Reviews," *2020 Int. Conf. Contemp. Comput. Appl. IC3A 2020*, no. May, pp. 217–220, 2020, doi: 10.1109/IC3A48958.2020.233300.

[9] A. Harun and D. Putri Ananda, "Analysis of Public Opinion Sentiment About Covid-19 Vaccination in Indonesia Using Naïve Bayes and Decission Tree," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 1, no. 1, pp. 58–64, 2021, doi: 10.57152/malcom.v1i1.63.

[10] É. R. Santana, L. Lopes, and R. M. de Moraes, "Recognition of the Effect of Vocal Exercises by Fuzzy Triangular Naive Bayes, a Machine Learning Classifier: A Preliminary Analysis," *J. Voice*, no. November, 2022, doi: 10.1016/j.jvoice.2022.10.001.

[11] A. Ali, A. Khairan, F. Tempola, and A. Fuad, "Application Of Naïve Bayes to Predict the Potential of Rain in Ternate City," *E3S Web Conf.*, vol. 328, p. 04011, 2021, doi: 10.1051/e3sconf/202132804011.

[12] D. P. Sari, D. Rosadi, A. R. Effendie, and Danardono, "Discretization methods for bayesian networks in the case of the earthquake," *Bull. Electr. Eng. Informatics*, vol. 10, no. 1, pp. 299–307, 2021, doi: 10.11591/eei.v10i1.2007.

[13] D. P. Sari, M. Rosha, and D. Rosadi, "Disaster Mitigation Efforts Using K-Medoids Algorithm and Bayesian Network," *EKSAKTA Berk. Ilm. Bid. MIPA*, vol. 23, no. 03, pp. 231–241, 2022, doi: 10.24036/eksakta/vol23-iss03/304.

[14] D. P. Sari, D. Rosadi, A. R. Effendie, and Danardono, "Application of Bayesian Network Model in Determining the Risk of Building Damage Caused by Earthquakes," *Int. Conf. Inf. Commun. Technol. ICOIACT*, pp. 131–135, 2018, doi: 10.1109/ICOIACT.2018.8350776.

[15] P. M. Singh and M. E. Languages, *Probability & Statistics*. 2020.

[16] D. P. Bertsekas and J. N. Tsitsiklis, *An Introduction to Probability*. 2011. doi: 10.1002/9781444390155.ch5.

[17] B. E. Putro and T. Saepurohman, "A Classification Approach to Predicting Beef Knuckle Quality using the Decision Tree and Naïves Bayes Method: Case Study: Tiga Bersaudara Factory," *2020 IEEE 7th Int. Conf. Ind. Eng. Appl. ICIEA 2020*, pp. 779–783, 2020, doi: 10.1109/ICIEA49774.2020.9102019.

[18] K. Komahan, "Naïve Bayes versus Support Vector Machines: A comparison of two algorithmic approaches for estimating cause of death from free-text," 2019.

[19] A. P. Dawid, "Conditional Independence in Statistical Theory," *J. R. Stat. Soc. Ser. B*, vol. 41, no. 1, pp. 1–15, 1979, doi: 10.1111/j.2517-6161.1979.tb01052.x.

[20] M. R. Romadhon and F. Kurniawan, "A Comparison of Naive Bayes Methods, Logistic Regression and KNN for Predicting Healing of Covid-19 Patients in Indonesia," *3rd 2021 East Indones. Conf. Comput. Inf. Technol. EIConCIT 2021*, pp. 41–44, 2021, doi: 10.1109/EIConCIT50028.2021.9431845.

[21] N. Salmi and Z. Rustam, "Naïve Bayes Classifier Models for Predicting the Colon Cancer," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 546, no. 5, 2019, doi: 10.1088/1757-899X/546/5/052068.

[22] M. J. Sánchez-Franco, A. Navarro-García, and F. J. Rondán-Cataluña, "A naive Bayes strategy for classifying customer satisfaction: A study based on online reviews of hospitality services," *J. Bus. Res.*, vol. 101, no. December, pp. 499–506, 2019, doi: 10.1016/j.jbusres.2018.12.051.

[23] T. A. Shaikh and R. Ali, "A CAD Tool for Breast Cancer Prediction using Naive Bayes Classifier," *2020 Int. Conf. Emerg. Smart Comput. Informatics, ESCI 2020*, pp. 351–356, 2020, doi: 10.1109/ESCI48226.2020.9167568.

[24] M. Umar, "SENTIMENT ANALYSIS OF STUDENTS' OPINION ON PROGRAMMING ASSESSMENT USING NAÏVE BAYES ALGORITHM ON SMALL DATA," *Prog. Retin. Eye Res.*, vol. 561, no. 3, pp. S2–S3, 2019.

[25] K. Yadav and R. Thareja, "Comparing the Performance of Naive Bayes And Decision Tree Classification Using R," *Int. J. Intell. Syst. Appl.*, vol. 11, no. 12, pp. 11–19, 2019, doi: 10.5815/ijisa.2019.12.02.