Comparison of Multinomial, Bernoulli, and Gaussian Naïve Bayes for Complaint Classification in Pro Denpasar Application

Ida Bagus Mahendra[1], I Made Gede Sunarya[2*], I Made Agus Wirawan[3]

[1,2,3]*Master of Computer Science Program, Universitas Pendidikan Ganesha, Indonesia*

*corr-author: sunarya@undiksha.ac.id

Abstract - *Pelayanan Rakyat Online Denpasar* or PRO Denpasar is a Denpasar City Government application intended as a public service mall to support Denpasar to become a smart city. This application was built since 2014 and is actively used in channeling public complaints so optimization is continuously needed to increase the efficiency of application use. Application optimization is carried out by developing decision support tools to determine complaint categories that are still done manually. The application of the proper artificial intelligence method can be used as a solution in classifying complaint categories to become a decision support tool for operators. This study compares three classification methods including multinomial naïve bayes, bernoulli naïve bayes and gaussian naïve bayes by applying TF-IDF feature extraction to determine the best complaint category classification method. Based on eight comparison scenario results by applying a comparison of 25%, 50%, 75% and 100% of complaint descriptions with 5-fold cross validation and 10-fold cross validation, it was found that the multinomial naïve Bayes method provided the best result in seven combined comparisons involving the test parameters accuracy, precision, recall, f1-score and processing time.

Keywords: Naïve Bayes; Natural Language Processing (NLP); Public Complaints; Text Mining; TF-IDF

I. INTRODUCTION

The rapid advancement of technology has ushered in a new era of digital transformation, affecting various sectors including business, healthcare, education, and government services. One prominent example of this transformation is the Pelayanan Rakyat Online Denpasar (PRO Denpasar) application, developed by the Denpasar City Government. PRO Denpasar aims to enhance public service delivery by providing a digital platform for citizens to submit complaints, inquiries, and suggestions. The development of this application has made Denpasar obtain the first predicate as an IKCI smart city in 2018 with an assessment using the smart city wheel indicator [1]. The application is actively used by the public. According to 2018 data, there were 1.693 complaints with each status, namely 1.132 had been followed up, 556 were in the initial response stage, and 6 complaints had not been followed up. In 2019, there was an increase in complaints reaching 1.851. Until 2020 there was 1.501 complaints [1]. Currently, the application plays an important role in channeling public complaints in Denpasar City. This makes optimizing the application a mandatory step to increase application efficiency. While PRO Denpasar has significantly improved accessibility to government services, the manual categorization of complaints remains a time-consuming and error-prone process. Automation of the selection of complaint categories can be one form of increasing the effectiveness and efficiency of the public complaint system process.

Research related to the classification of complaint categories has previously been conducted [2], [3], [4], using various data sources, different number of classes with the naïve bayes method. The test results state that this method is able to produce good accuracy values with an average score above 80%. In the case of a comparison of complaint report dataset classification methods [5], the naïve bayes method is able to outperform other methods including Random Forest (RF), Support Vector Machine (SVM) and K-Nearest Neighbors (KNN) by obtaining the highest accuracy at 80%. Research [6] also discusses the comparison of complaint category classification using the RF, SVM, and naïve bayes methods but different evaluation results were found. From the two test scenarios without using a sampling model and using a sampling model, it was found that the naïve bayes method obtained the lowest results in terms of accuracy, precision, recall, and f1-score. Despite obtaining the lowest evaluation results, the naïve bayes method was the most superior in terms of processing time, which is also an important aspect of system effectiveness. The results of this comparison provide an opportunity to improve the performance of the naïve bayes method while maintaining its advantages, being insensitive to imbalanced data, efficient computing resources and fast processing time [6], [7], [8]. In an effort to improve performance, the implementation of naïve bayes method variance can be an option. In the study [9] which proposed a naïve bayes and n-gram multinomial variant model for the classification of online complaint documents, it was able to provide the highest 88.23% accuracy results. The bernoulli naïve bayes variance applied to the study of English online news hoax detection with TF-IDF feature extraction [10] produced a prediction model with an accuracy value of 98.5% on news data, surpassing previous research with an accuracy increase of 16.08%. Comparison of multinomial naïve bayes and bernoulli naïve bayes in text document classification found the highest accuracy in bernoulli naïve bayes with a value of 71.33% [11].

In this study, we have conducted a comparative analysis of three Naïve Bayes variants of multinomial naïve bayes, bernoulli naïve bayes, and gaussian naïve bayes. By evaluating their performance on a dataset of public complaints from PRO Denpasar, we aim to identify the optimal variant for automating the categorization process. The research is expected to make it easier for operators to determine complaint categories by acting as a decision-making tool.

II. METHOD

The research design is arranged in several sequential steps, namely from inputting public complaint data, data pre-processing, separation for training data and test data, data feature extraction, data classification, evaluation and output in the form of comparative analysis results. In general, the research process described by the Fig. 1.

A. Research Data Overview

The research data was obtained from secondary data sources, namely public complaint data from PRO Denpasar application specifically the Denpasar Communication, Informatics and Statistics Office through a written process accommodated by the administration section of Ganesha University. The research sample used is complaint data from the Pelayanan Rakyat Online Denpasar (PRO Denpasar) application in 2014-2023. In this study, the data used were complaint descriptions and complaint categories. The complaint descriptions are used as the main data that will be processed to obtain complaint category class classifications, meanwhile the complaint category is used as the ground truth of the classification class used. There are 6 categories obtained and confirmed directly by the office to be used as research classes, namely the categories "*Apresiasi*", "*Informasi*", "*Keluhan*", " *Permohonan Informasi Publik (PPID)*", "*Pertanyaan*", and "*Usul / Saran*". These categories align with the contents of Article 3 of Denpasar Mayor Regulation number 45 of 2013, which outlines complaint handling services in Denpasar City [12] and is employed by the Denpasar City Communication, Informatics, and Statistics Office. A total of 15.147 data points were provided and utilized in this research. To simplify the explanation process, only 12 documents will be displayed, with 10 for training and 2 (documents 11 and 12) for testing. Table I presents examples of the provided data.

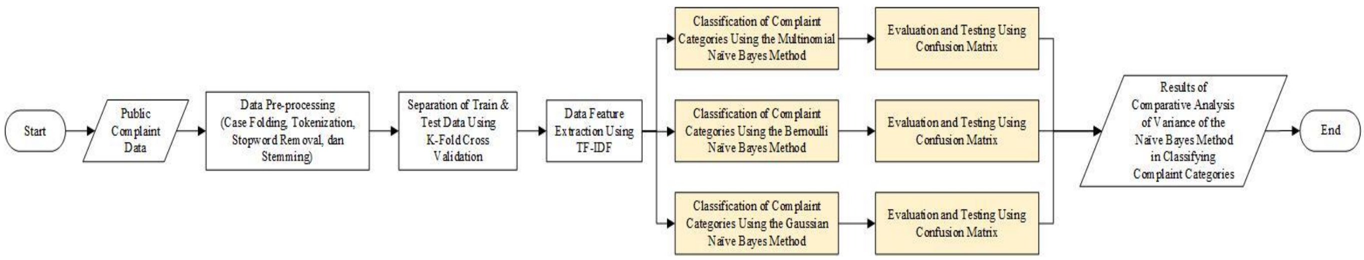

Fig. 1 Research design flowchart

TABLE I
COMPLAINT SAMPLE DATASET

No	Complaint Description	Class
1	* Pelayanan di dinas kependudukan dan pencatatan sipil di bagian validasi sangat memuaskan. *	*Apresiasi*
…	…………………………………………	…………..
12	* Mohon informasi Ringkasan LKPD Kota Denpasar TA 2017 dan 2018 Penelitian. Terimakasih *	*Permohonan Informasi Publik (PPID)*

B. Data Pre-Processing

Research data processing are separated into 4 main process which are case folding, tokenization, stopword removal and stemming. Case folding is the process of changing characters into the same form [12], [13]. This process is done by changing all letters to capital or lower case. Table II presents the examples of case folding.

Tokenization is a process of identifying individual words in a sentence called tokens [12], [13]. The process includes removing some parts such as correcting abbreviations, removing numbers, punctuation, spaces and other characters that are considered to have no effect on text processing [14]. Table III presents the example of tokenization.

Stopword removal is known as a stage of selecting words that are considered important so that words that have no meaning can be removed [12], [13]. The main goal in implementing this stopwords process is to reduce the number of words in a document. Table IV presents the examples of stopword removal.

Stemming is the process of reducing words in a document to their root forms using language-specific rules [12], [13]. Table V presents the example of stemming.

C. Data Separation Process

The separation of training data and test data in this study was carried out by applying the k-fold cross validation method and feature separation based on percentage. The comparison of 5-fold and 10-fold cross validation values used in this study follows a number of references which state that these values are suitable to use for better result. The 5-fold cross validation value is used in the study [3], [8], [15], [16] while the 10-fold cross validation value is used in the study [2], [10], [15], [17], [18], [19], [20]. Each block process is carried out for each feature separation of 25%, 50%, 75% and 100% respectively to see the comparison with the number of features [21], [22], [23]. Fig. 2 illustrates the data separation process.

TABLE II
CASE FOLDING STAGES EXAMPLE

No	Data Before Process	Data After Process
1	* Pelayanan di dinas kependudukan dan pencatatan sipil di bagian validasi sangat memuaskan. *	* pelayanan di dinas kependudukan dan pencatatan sipil di bagian validasi sangat memuaskan. *
…	…………………	………………
12	* Mohon informasi Ringkasan LKPD Kota Denpasar TA 2017 dan 2018 Penelitian. Terimakasih *	* mohon informasi ringkasan lkpd kota denpasar ta 2017 dan 2018 penelitian. terimakasih *

TABLE III
EXAMPLE OF TOKENIZATION STAGES

No	Data Before Process	Data After Process
1	* Pelayanan di dinas kependudukan dan pencatatan sipil di bagian validasi sangat memuaskan. *	*'pelayanan', 'di', 'dinas', 'kependudukan', 'dan', 'pencatatan', 'sipil', 'di', 'bagian', 'validasi', 'sangat', 'memuaskan'*
…	………………	………………
12	* mohon informasi ringkasan lkpd kota denpasar ta 2017 dan 2018 penelitian. terimakasih *	*'mohon', 'informasi', 'ringkasan', 'lkpd', 'kota', 'denpasar', 'ta', 'dan', 'penelitian', 'terimakasih'*

TABLE IV
EXAMPLE OF STOPWORD REMOVAL STAGES

No	Data Before Process	Data After Process
1	*'pelayanan', 'di', 'dinas', 'kependudukan', 'dan', 'pencatatan', 'sipil', 'di', 'bagian', 'validasi', 'sangat', 'memuaskan'*	*'pelayanan', 'dinas', 'kependudukan', 'pencatatan', 'sipil', 'validasi', 'memuaskan'*
…	……………	……………
12	*'mohon', 'informasi', 'ringkasan', 'lkpd', 'kota', 'denpasar', 'ta', 'dan', 'penelitian', 'terimakasih'*	*'mohon', 'informasi', 'ringkasan', 'lkpd', 'kota', 'denpasar', 'ta', 'penelitian', 'terima kasih'*

TABLE V
EXAMPLE OF STEMMING STAGES

No	Data Before Process	Data After Process
1	*'pelayanan', 'dinas', 'kependudukan', 'pencatatan', 'sipil', 'validasi', 'memuaskan'*	*'layan', 'dinas', 'duduk', 'catat', 'sipil', 'validasi', 'muas'*
…	……………	…………
12	*'mohon', 'informasi', 'ringkasan', 'lkpd', 'kota', 'denpasar', 'ta', 'penelitian', 'terima kasih'*	*'mohon', 'informasi', 'ringkas', 'lkpd', 'kota', 'denpasar', 'ta', 'teliti', 'terima', 'kasih'*

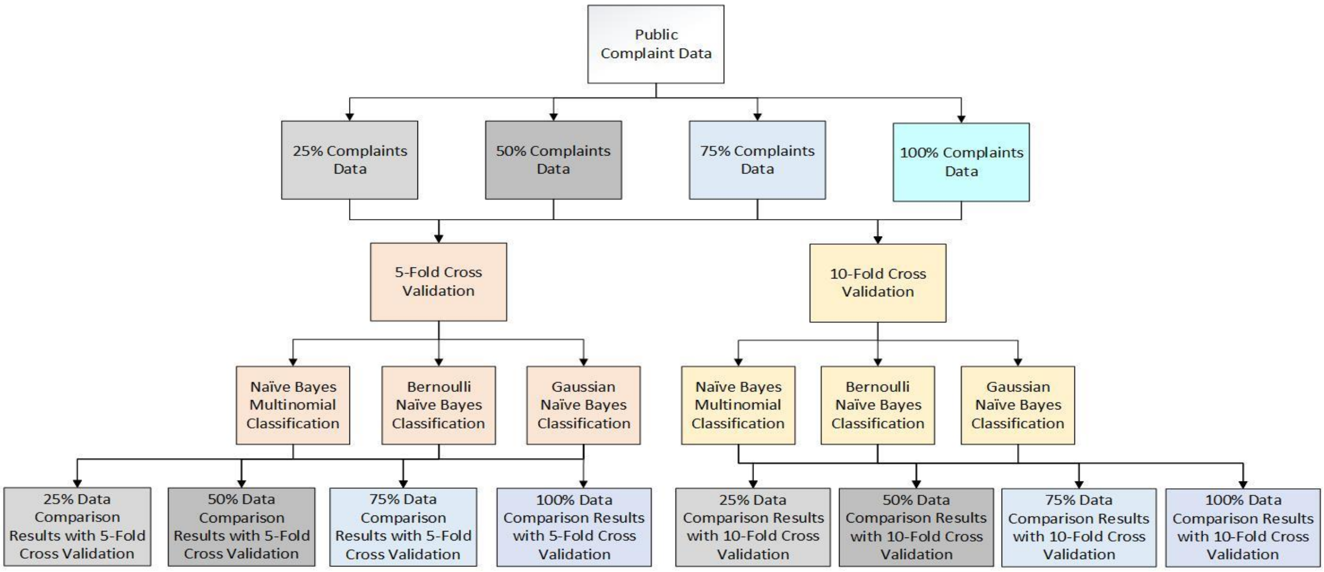

Fig. 2 Separation of training data and test data

D. Feature Extraction

Feature extraction is the stage of obtaining features based on pre-processing data to be used in classification. Feature extraction in this study uses Term Frequency-Inverse Document Frequency (TF-IDF). This method calculates how many times different words (terms) appear in a collection of documents and maps several levels of importance from the resulting frequencies [2], [14]. The formula of TF-IDF uses (1) [24].

$$tfidf(w) = tf \times log\frac{n}{df(w')} \quad (1)$$

TF-IDF calculations are carried out on training data for each iteration of fold cross validation. Table VI presents the examples of TF-IDF feature extraction.

E. Feature Classification

Feature classification is a stage for training machine learning based on known data to determine the type of new unknown data [24]. The classification process involves processing the classification method on the weight data from the TF-IDF feature extraction stage. Based on the naïve Bayes theory, an equation is stated as (2) [25].

$$P(H|X) = \frac{P(X|H)\ X\ P(H)}{P(X)} \quad (2)$$

The prior value shows the initial probability when H occurs [25]. Given A is a word or feature, C is a value or class category then the formula can be written as (3) [8].

$$P(A) = \frac{Nc}{N} \quad (3)$$

Likelihoods shows the probability that the occurrence of event X will affect H [25]. Likelihood is processed from the (4).

$$P(C|A) = \frac{count(C_i A)}{\sum_{C \epsilon V}(count(c,a))} \quad (4)$$

Multinomial Naïve Bayes (MNB model uses a multinomial distribution, where features can be obtained from discrete values such as the number of words [20]. In Multinomial Naive Bayes, likelihoods are obtained by summing the TF-IDF, adding a Laplace smoothing constant of 1, and dividing the sum by total of TF-IDF documents [26]. Table VII presents the example process of calculating likelihood.

TABLE VI
EXAMPLE OF TF-IDF CALCULATION RESULTS

No	Word	TF -IDF									
		DOC 1	DOC 2	DOC 3	DOC 4	DOC 5	DOC 6	DOC 7	DOC 8	DOC 9	DOC 10
1	*action*	0	0	0	0	0	0,1	0	0	0	0
…	………	……….	………	………	………	………	………	………	………	………	………
71	*validasi*	0,3289	0	0	0	0	0	0	0	0	0

TABLE VII
NAÏVE BAYES ALGORITHM LIKELIHOODS EXAMPLE

No	Word	P(Word$_n$ |Apresiasi)	P(Word$_n$ | Informasi)	P(Word$_n$ |Keluhan)	P(Word$_n$ |PPID)	P(Word$_n$ |Pertanyaan)	P(Word$_n$ |Usul/Saran)
1	*action*	0,01329879	0,01334339	0,016398491	0,01336211	0,013642118	0,01366059
…	………	…………..	…………….	…………..	…………..	……………..	………………
71	*validasi*	0,01767276	0,01334339	0,013329939	0,01336211	0,013642118	0,01366059

The posterior value or classification of document 11 and 12 then can be calculated for each class using simplified formula as (5) [25].

$$P(C|A1, A2, \ldots, A_n) = \arg\max C \left(\prod_{i=1}^{n} \left(P(A_i|C) X\, P(C)\right)\right) \tag{5}$$

Bernoulli Naïve Bayes (BNB) emphasizes the presence or absence of a term in the document under consideration [20]. BNB method requires binary data to be processed so that the TF-IDF feature needs to be binarized into the number 0 or 1 by setting a threshold. Table VIII provides the examples of processed data.

The calculation of the likelihood value in BNB calculation can be written as (6) and (7) [27].

$$P(X_i|y) = P(i|y)X_i + \left(1 - P(i|y)\right)(1 - X_i) \tag{6}$$

$$P(i|y) = {N_{i,y} + 1}\big/{N_y + 2} \tag{7}$$

Gaussian Naïve Bayes (GNB) method is well known designed to handle text data with continuous distribution features, such as numeric values or continuous attributes [20]. The probability P(X|H) is calculated by utilizing the Gaussian probability density function as (8) [20], [28].

$$P(X|H) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e - \frac{x-\mu^2}{2\sigma^2} \tag{8}$$

The mean and standard deviation (variance) values are required to calculate gaussian probability density function. Below are the formulas of mean and variance (9) and (10).

$$\mu = \frac{\Sigma_{i=1}^{n} xi}{n} \tag{9}$$

$$\sigma = \sqrt{\frac{\Sigma_{i=1}^{n}(xi-\mu)^2}{n}} \tag{10}$$

F. Model Evaluation

Evaluation is intended to measure and analyze the results of the classification process and the processing time required. Evaluation process use the confusion matrix table that describes the performance of a specific model or algorithm [29]. Parameters are used namely accuracy, precision, recall, and f1-score from the classification process carried out. Data can be used to measure the performance of a model including, accuracy or overall total how often the model correctly classifies.

Given the confusion matrix and time process data, normalization is then used to scale data for quantitative studies [30]. The cost benefit max-min normalization method is used as form of data normalization intended to determine the normal value of matrix data by utilizing assessments based on costs (losses) and benefits (advantages). The general formula is as (11) [30], [31], [32], [33].

$$n_{ij} = \begin{cases} \frac{Min_i(x_{ij})}{x_{ij}}, & cost \\ \frac{x_{ij}}{Max_i(x_{ij})}, & benefit \end{cases} \tag{11}$$

TABLE VIII
FEATURE BINARY CALCULATION EXAMPLE

No	Word	TF -IDF									
		DOC 1	DOC 2	DOC 3	DOC 4	DOC 5	DOC 6	DOC 7	DOC 8	DOC 9	DOC 10
1	*action*	0	0	0	0	0	1	0	0	0	0
…	………	……….	………	………	………	………	………	………	………	………	………
71	*validasi*	1	0	0	0	0	0	0	0	0	0

III. RESULT AND DISCUSSION

A. Data Exploration

Data description is needed to understand the characteristics of research data through the description of the characteristics and forms of the data to obtain a specific picture of the conditions of the research data used. The description of this research data follows reference [34] and includes several characteristics presented in Table IX.

Characteristics information provide some conclusions that can be drawn, like the average document length indicates that in general the documents in this dataset are of moderate size. The average word length of only 6 words suggests that the words used tend to be short, while very high total of unique words may indicate that this document collection has a very extensive vocabulary. The total words reaching hundreds of thousands indicate that the volume of data analyzed is quite large provides a strong basis for conducting a more in-depth analysis. Consistently appearing words can be visualized using a WordCloud, as illustrated in Fig. 3.

The analysis results indicate that some of the most frequently appearing words are jalan, data, and nama. This information could be very helpful in filtering words, or to get a sense of the overall content of a document or corpus.

TABLE IX
RESEARCH DATA CHARACTERISTICS

No	Data Characteristics	Value
1	Average Document Length	49
2	Average Word Length	6
3	Total Unique Words	67,241
4	Total Words	741,694
5	Longest Word Count in Document	1,863
6	Shortest Word Count in Document	1

B. Classification Methods Results Comparison

Comparative evaluation of multinomial naïve bayes, bernoulli naïve bayes and gaussian naïve bayes classification methods will be carried out in two types. First, the comparative evaluation between the number of blocks or fold values in cross validation which in this study uses the values k = 5 and k = 10. The next form of comparison involves separating the amount of data into several parts, namely 25%, 50%, 75 and 100%. The comparative evaluation uses average value for each metrics which are shown at Table X.

In 50% data scenario shown in Table XI, the 5-fold and 10-fold cross validation comparison shows similar results where the multinomial naïve bayes method has the best evaluation value for accuracy, recall and processing time metrics, and the bernoulli naïve bayes method for precision and f1-score metrics. Table XII shows the classification methods comparison using 75% data, while using 100% data presented in Table XIII.

Fig. 3 Word cloud results for total complaint data

TABLE X
CLASSIFICATION METHODS COMPARISON USING 25% DATA WITH 5 AND 10-FOLD CROSS VALIDATION

No	Methods	5-Fold Cross Validation					10-Fold Cross Validation				
		Accuracy	Precision	Recall	F1 - Score	Process Time	Accuracy	Precision	Recall	F1 - Score	Process Time
1	MNB	**0.7118**	0.5081	**0.7118**	0.5925	0.0082	**0.7118**	0.5094	**0.7118**	0.5931	**0.0071**
2	BNB	0.7004	0.5988	0.7004	**0.6150**	**0.0080**	0.7015	**0.5983**	0.7015	**0.6187**	0.0085
3	GNB	0.6441	**0.5997**	0.6441	0.6090	0.7384	0.6325	0.5970	0.6325	0.6031	0.6851

TABLE XI
CLASSIFICATION METHODS COMPARISON USING 50% DATA WITH 5 AND 10-FOLD CROSS VALIDATION

No	Methods	5-Fold Cross Validation					10-Fold Cross Validation				
		Accuracy	Precision	Recall	F1 - Score	Process Time	Accuracy	Precision	Recall	F1 - Score	Process Time
1	MNB	**0.7455**	0.5576	**0.7455**	0.6375	**0.01308**	**0.7455**	0.5583	**0.7455**	0.6377	**0.0152**
2	BNB	0.7377	**0.6448**	0.7377	**0.6713**	0.0169	0.7391	**0.6513**	0.7391	**0.6752**	0.0186
3	GNB	0.6222	0.6301	0.6222	0.6226	2.94138	0.6126	0.6302	0.6126	0.6175	3.0899

TABLE XII
CLASSIFICATION METHODS COMPARISON USING 75% DATA WITH 5 AND 10-FOLD CROSS VALIDATION

No	Methods	5-Fold Cross Validation					10-Fold Cross Validation				
		Accuracy	Precision	Recall	F1 - Score	Process Time	Accuracy	Precision	Recall	F1 - Score	Process Time
1	MNB	**0.7770**	0.6078	**0.7770**	0.6809	**0.02018**	**0.7770**	0.6088	**0.7770**	0.6813	**0.0214**
2	BNB	0.7622	**0.6742**	0.7622	**0.7024**	0.02548	0.7639	**0.6782**	0.7639	**0.7059**	0.0295
3	GNB	0.6138	0.6603	0.6138	0.6309	6.81214	0.6014	0.6622	0.6014	0.6254	6.9554

Judging from the average value of 5-fold and 10-fold cross validation for 75% data shown in Table XII, it is concluded that the multinomial naïve bayes method has the best evaluation value for the accuracy, recall and processing time metrics, and the bernoulli naïve bayes method for the precision and f1-score metrics. As shown in Table XIII, the 5-fold and 10-fold cross validation scenario also shows similar results with the 50% data and 75% data that the multinomial naïve bayes method still has the best evaluation value for the accuracy, recall and processing time metrics, while the bernoulli naïve bayes method for the precision and f1-score metrics.

These findings demonstrate that different variants of classification models exhibit varying results across different scenarios. Among the eight scenarios analyzed, multinomial naive bayes generally achieved the highest accuracy, recall and process time, while bernoulli naive bayes demonstrated superior performance in terms of precision and f1-score. Multinomial naive bayes effectively captures the nuances of language by considering word frequencies, leading to improved accuracy and recall [34], [35]. In contrast, bernoulli naive bayes, by disregarding word frequencies, may be less susceptible to overfitting on frequent words that are not truly indicative of a specific category, resulting in better precision [34], [36]. The f1-score, which balances precision and recall, indicates that bernoulli naive bayes achieves a good balance between these two metrics. Multinomial naive bayes is also generally faster to train and predict, especially with larger datasets [36]. Research suggests that bernoulli naive bayes may be more suitable for short sentences [36], which could explain why it initially exhibited the best time processing. Furthermore, research [34] indicates that gaussian naive bayes may perform better with smaller datasets, aligning with the initial observation where gaussian naive bayes demonstrated the best recall with 25% data and 5-fold cross-validation before being surpassed by bernoulli naive bayes in all subsequent scenarios.

C. *Determining the Best Model from Results Evaluation*

This section presents a comparison of the results of the cost-benefit max-min normalization calculation against all methods tested (Table XIV). The discussion is as follows.

TABLE XIII
CLASSIFICATION METHODS COMPARISON USING 100% DATA WITH 5 AND 10-FOLD CROSS VALIDATION

No	Methods	5-Fold Cross Validation					10-Fold Cross Validation				
		Accuracy	Precision	Recall	F1 - Score	Process Time	Accuracy	Precision	Recall	F1 - Score	Process Time
1	MNB	**0.8057**	0.6549	**0.8057**	0.7210	**0.0266**	**0.8057**	0.6709	**0.8057**	0.7298	**0.02952**
2	BNB	0.7837	**0.7012**	0.7837	**0.7289**	0.03222	0.7848	**0.7051**	0.7848	**0.7323**	0.03639
3	GNB	0.6264	0.7007	0.6264	0.6529	11.41718	0.6153	0.7031	0.6153	0.6488	12.05507

TABLE XIV
COST-BENEFIT MAX-MIN NORMALIZATION CALCULATION RESULTS

No	Classification Methods	K-Fold Cross Validation	25% Data	50% Data	75% Data	100% Data
1	Multinomial Naïve Bayes	5-Fold	0.9572894	**0.9628805**	**0.9741815**	**0.9846157**
		10-Fold	**0.9620074**	**0.9603408**	**0.9725668**	**0.9896230**
2	Bernoulli Naïve Bayes	5-Fold	**0.9933114**	0.9506079	0.9507905	0.9541929
		10-Fold	0.9600941	0.9596748	0.9380689	0.9518469
3	Gaussian Naïve Bayes	5-Fold	0.7622346	0.7157050	0.6920765	0.6904628
		10-Fold	0.7520597	0.7061485	0.6826531	0.6825750

As depicted in Table XIV, the analysis results found that the multinomial naïve bayes method is the best method to use in classifying the complaint categories of the Pro Denpasar application. This determination is supported by eight different types of calculation scenarios using the cost-benefit max-min normalization method for each classification method. The multinomial naïve bayes method obtained the highest results seven times out of eight trials, where in one trial the highest value was obtained by the bernoulli naïve bayes method. The evaluation results show that naïve bayes is a good text classification method for complaint report datasets, as proven by the highest accuracy of 0.80574 achieved by multinomial naïve bayes. This claim is also supported by earlier research that achieved similar results metrics [2], [3], [4]. It was also found that naïve bayes variance can influence the results, as notably mentioned by research [9], [10], [11]. These evaluation results are mostly obtained because the characteristics of the multinomial naïve Bayes model assume that discrete count data, or the number of each word, is considered important and plays a significant role in the entire calculation of the multinomial distribution [34], [35]. This allows unique words to emphasize their value. Additionally, TF-IDF feature extraction effectively calculates the importance of words in a document relative to the corpus, further helping to emphasize the value of unique words [37]. Given the average document length of 49 words, this also contribute to several implications as why bernoulli naive bayes might not be well-suited for long sentences [36], which are likely prevalent in the complaint classification dataset. Furthermore, gaussian naive bayes may not perform optimally with larger datasets [34], as observed in the complaint classification dataset.

IV. CONCLUSION

This study compared three classification methods of multinomial naïve bayes, bernoulli naïve bayes and gaussian naïve bayes to obtain the most appropriate naïve bayes method used in classifying complaint categories in the PRO Denpasar application assessed from a combination of evaluations of accuracy, precision, recall, and the highest f1-score and the lowest processing time. Based on the entire analysis and evaluation process through eight experimental scenarios, it was found that the multinomial naïve bayes method obtained the highest results from seven experiments through a comparison of the combined evaluation from the calculation results of cost benefit max-min normalization. Future research on complaint data classification could focus on improving the evaluation results of the Gaussian Naive Bayes method. This study revealed that Gaussian Naive Bayes yielded the lowest evaluation scores. Therefore, a series of combined methods or different experiments could be conducted to achieve better classification results and address the limitations of this research. Other potential research topics include classification for prioritizing complaints or developing additional support systems to streamline and improve the efficiency of complaint handling processes for staff.

ACKNOWLEDGEMENT

The authors sincerely appreciate everyone who has supported and contributed to this research. We are grateful for the valuable advice, constructive criticism, and insightful input provided throughout the study.

REFERENCES

[1] A. D. Syafitri, E. Purnamaningsih, and P. D. Yudartha, “Pengaruh Kualitas Pelayanan Terhadap Kepuasan Penggunapada Aplikasi Pro-Denpasar di Kota Denpasar,” *Jurnal Administrasi Publik Citizen Charter*, vol. 2, no. 1, Jun. 2022.

[2] H. A. Gozali and M. A. Rosid, “Classification of Student Complaints with Naive Bayes and Literary Methods Klasifikasi Keluhan Mahasiswa dengan Metode Naive Bayes dan Sastrawi,” 2020.

[3] C. Anam, “Klasifikasi Pengaduan Pelayanan Dispendukcapil Kota Malang Menggunakan Metode Naïve Bayes dan Seleksi Fitur Glasgow-II,” *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 4, no. 9, pp. 3264–3271, Sep. 2020, [Online]. Available: http://j-ptiik.ub.ac.id

[4] F. A. Lestari, L. Efrizoni, E. Ali, and R. Rahmiati, “Sistem Klasifikasi Pengaduan Masyarakat Pada BPJS Ketenagakerjaan Menggunakan Algoritma Naïve Bayes Berbasis Mobile,” *Building of Informatics, Technology and Science (BITS)*, vol. 4, no. 1, Jun. 2022, doi: 10.47065/bits.v4i1.1685.

[5] G. Angiani, M. Mordonini, U. Pietroni, P. Fornacciari, G. Lombardo, and M. Tomaiuolo, “Automatic processing and classification of citizens’ reports,” in *ACM International Conference Proceeding Series*, Association for Computing Machinery, Nov. 2018, pp. 310–311. doi: 10.1145/3284869.3284930.

[6] I. G. N. A. Kusuma, I. M. Pradipta, I. M. A. Santosa, and I. K. Dharmendra, “Penanganan Ketidakseimbangan Data pada Klasifikasi Pengaduan Masyarakat,” *Jurnal Teknologi Informasi dan Komputer*, vol. 9, no. 5, Oct. 2023, [Online]. Available: https://pengaduan.denpasarkota.go.id[6].

[7] J. P. D, Delizo, “Philippine Twitter Sentiments during Covid-19 Pandemic using Multinomial Naïve-Bayes,” *International Journal of Advanced Trends in Computer*

Science and Engineering, vol. 9, no. 1.3, pp. 408–412, Jun. 2020, doi: 10.30534/ijatcse/2020/6491.32020.

[8] A. Sabrani, I. W. G. P. W. Wedashwara, and F. Bimantoro, “Metode Multinomial Naïve Bayes untuk Klasifikasi Artikel Online Tentang Gempa di Indonesia,” *JTIKA*, vol. 2, no. 1, Mar. 2020, [Online]. Available: http://jtika.if.unram.ac.id/index.php/JTIKA/

[9] F. A. Saputra and C. Dewi, “Klasifikasi Dokumen Pengaduan Sambat Online menggunakan Metode Multinomial Naïve Bayes dan N-Gram,” 2020. [Online]. Available: http://j-ptiik.ub.ac.id

[10] A. Y. Prayoga, A. Id Hadiana, and F. R. Umbara, “Deteksi Hoax pada Berita Online Bahasa Inggris Menggunakan Bernoulli Naïve Bayes dengan Ekstraksi Fitur Tf-Idf,” *Jurnal Syntax Admiration*, vol. 2, no. 10, pp. 1808–1823, Oct. 2021, doi: 10.46799/jsa.v2i10.327.

[11] Widyawati and Sutanto, “Perbandingan Kinerja Variasi Naïve Bayes Multivariate Bernoulli dan Naïve Bayes Multinomial dalam Pengklasifikasian Dokumen Teks,” *JURNAL OF INNOVATION AND FUTURE TECHNOLOGY (IFTECH)*, vol. 2, no. 1, Feb. 2020.

[12] K. V Deshmukh and S. S. Shiravale, “Ambiguity Resolution in English Language for Sentiment Analysis,” in *IEEE Punecon*, Pune, 2018.

[13] H. Zhang, J. Nie, and H. Wang, “Product Perspective: A Study of User Demand Priorities in Online Community Reviews,” 2019.

[14] N. M. Ranjan and R. S. Prasad, “Text Analytics: An Application of Text Mining,” *Journal of Data Mining and Management*, vol. 6, no. 3, pp. 1–6, Nov. 2021, doi: 10.46610/jodmm.2021.v06i03.001.

[15] E. K. Andana, M. Othman, and R. Ibrahim, “Comparative Analysis of Text Classification Using Naive Bayes and Support Vector Machine in Detecting Negative Content in Indonesian Twitter,” *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 8, no. 1.3 S1, pp. 356–362, 2019, doi: 10.30534/ijatcse/2019/6481.32019.

[16] R. Fajar, “Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter,” *JURNAL INOVTEK POLBENG*, vol. 3, no. 1, Jun. 2018.

[17] A. D. Adyatma, L. Afuan, and E. Maryanto, “The Effect of Unigram and Bigram in the Naïve Bayes Multinomial for Analyzing of Comment Sentiment of Gojek Application in Google Play Store,” *Jurnal Teknik Informatika (JUTIF)*, vol. 4, no. 6, pp. 1535–1540, Dec. 2023, doi: 10.52436/1.jutif.2023.4.6.1310.

[18] K. S. Nugroho, I. Istiadi, and F. Marisa, “Naive Bayes classifier optimization for text classification on e-government using particle swarm optimization,” *Jurnal Teknologi dan Sistem Komputer*, vol. 8, no. 1, pp. 21–26, Jan. 2020, doi: 10.14710/jtsiskom.8.1.2020.21-26.

[19] M. A. Suprayogi, “Klasifikasi Pengaduan Laras Online Berbasis Text Mining Menggunakan Algoritma K-Nearest Neighbor dan Naïve Bayes,” *Jurnal Ilmiah Informatika Komputer*, vol. 26, no. 1, pp. 65–77, 2021, doi: 10.35760/ik.2021.v26i1.3397.

[20] T. Taslim, S. Handayani, and F. Fajrizal, “Kinerja Komparatif Optimasi Algoritma Naive Bayes dalam Klasifikasi Teks untuk Uji Klinis Kanker,” *Jurnal Eksplora Informatika*, vol. 13, no. 1, pp. 113–123, Sep. 2023, doi: 10.30864/eksplora.v13i1.994.

[21] D. Aziz, N. Hayatin, C. Sri, and K. Aditya, “Perbandingan Algoritma Selection Feature Query Expansion Ranking dan Symmetrical Uncertainty Pada Instagram Terkait Sengketa ZEE di Perairan Natuna,” *REPOSITOR*, vol. 3, no. 4, pp. 411–418, Aug. 2021.

[22] D. Maulana and R. S. S. Markiano, “Klasifikasi Analisis Sentimen Terhadap Calon Presiden 2019 pada Media Sosial Twitter Menggunakan Metode Algoritma Naïve Bayes,” *Jurnal Teknologi Pelita Bangsa*, vol. 11, no. 3, pp. 2407–3903, Mar. 2020.

[23] A. Y. Permana and D. N. Mahardhika, “Klasifikasi Analisis Sentimen Terhadap Kasus Penusukan Wiranto Media Sosial Twitter dengan Menggunakan Metode Naïve Bayes Classification,” *Jurnal Teknologi Pelita Bangsa*, vol. 10, no. 1, Mar. 2019.

[24] A. Lasarudin, “Klasifikasi Pengaduan Masyarakat Menggunakan Naive Bayes Berbasis Seleksi Atribut Information Gain,” 2018. [Online]. Available: http://research.

[25] M. B. Ressan and R. F. Hassan, “Naïve-Bayes family for sentiment analysis during COVID-19 pandemic and classification tweets,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 28, no. 1, pp. 375–383, Oct. 2022, doi: 10.11591/ijeecs.v28.i1.pp375-383.

[26] A. Nugraha and M. U. Siregar, “Application of The Naive Bayes Classifier Method In The Sentiment Analysis of Twitter User About The Capital City Relocation,” *PROC. INTERNAT. CONF. SCI. ENGIN*, vol. 4, Feb. 2021.

[27] W. M. P. Dhuhita and F. Zone, “Perbandingan Kinerja Algoritma Multinomial dan Bernoulli Naïve Bayes dalam Mengklasifikasikan Komentar Cyberbullying,” *Komputika : Jurnal Sistem Komputer*, vol. 12, no. 2, pp. 109–117, Oct. 2023, doi: 10.34010/komputika.v12i2.9767.

[28] N. A’ayunnisa, Y. Salim, and H. Azis, “Analisis performa metode Gaussian Naïve Bayes untuk klasifikasi citra tulisan tangan karakter arab,” *Indonesian Journal of Data and Science (IJODAS)*, vol. 3, no. 3, pp. 115–121, 2022.

[29] I. W. Saputro and B. W. Sari, “Uji Performa Algoritma Naïve Bayes untuk Prediksi Masa Studi Mahasiswa Naïve Bayes Algorithm Performance Test for Student Study Prediction,” *Citec Journal*, vol. 6, no. 1, 2019.

[30] A. Aytekin, “Comparative analysis of normalization techniques in the context of MCDM problems,”

Decision Making: Applications in Management and Engineering, vol. 4, no. 2, pp. 1–25, 2021, doi: 10.31181/dmame210402001a.

[31] Y. A. Singgalen, "Analisis Sentimen dan Sistem Pendukung Keputusan Menginap di Hotel Menggunakan Metode CRISP-DM dan SAW," *Journal of Information System Research (JOSH)*, vol. 4, no. 4, pp. 1343–1353, Jul. 2023, doi: 10.47065/josh.v4i4.3917.

[32] Y. Susanti, Y. Diah, and D. D. Hanum, "Penerapan Metode Naive Bayes dan Simple Additive Weighting dalam Penentuan Penerima Program Keluarga Harapan," *Seminar Nasional Sains dan Teknologi Terapan*, vol. VI, 2018.

[33] I. M. A. Wirawan, R. Wardoyo, D. Lelono, and S. Kusrohmaniah, "Continuous Capsule Network Method for Improving Electroencephalogram-Based Emotion Recognition," *Emerging Science Journal*, vol. 7, no. 1, pp. 116–134, Feb. 2023, doi: 10.28991/ESJ-2023-07-01-09.

[34] C. Naulak and P. Jindal, "A comparative study of Naive Bayes Classifiers with improved technique on Text Classification," *TechRxiv*, May 2022.

[35] T. Simbolon, A. P. Wibawa, I. A. E. Zaeni, and A. R. Ismail, "Text classification of traditional and national songs using naïve bayes algorithm," *Science in Information Technology Letters*, vol. 3, no. 2, pp. 59–72, Nov. 2022, doi: 10.31763/sitech.v3i2.1215.

[36] M. Abbas, K. Ali, A. Jamali, K. Ali Memon, and A. Aleem Jamali, "Multinomial Naive Bayes Classification Model for Sentiment Analysis," *IJCSNS International Journal of Computer Science and Network Security*, vol. 19, no. 3, p. 62, 2019, doi: 10.13140/RG.2.2.30021.40169.

[37] S. M. Sadid, J. C. Young, and A. Rusli, "Spam Filtering on User Feedback via Text Classification using Multinomial Naïve Bayes and TF-IDF," *Ultimatics : Jurnal Teknik Informatika*, vol. 13, no. 2, p. 108, 2021.