2021 3rd East Indonesia Conference on Computer and Information Technology (EIConCIT)
April 09-11. 2021, ISTTS Surabaya, Indonesia

Categorization of Exam Questions based on Bloom Taxonomy using Naïve Bayes and Laplace Smoothing

2021 3rd East Indonesia Conference on Computer and Information Technology (EIConCIT) | 978-1-6654-0514-0/20/$31.00 ©2021 IEEE | DOI: 10.1109/EIConCIT50028.2021.9431862

1st Eka Rahayu Setyaningsih
Teknologi Informasi
Insitut Sains dan Teknologi Terpadu Surabaya (ISTTS)
Surabaya, Indonesia
eka@stts.edu

2nd Indah Listiowarni
Teknologi Informasi
Insitut Sains dan Teknologi Terpadu Surabaya (ISTTS)
Surabaya, Indonesia
listioindah@gmail.com

***Abstract* — Being famous for a classification algorithm using a simple statistic calculation, Naive Bayes produces a relatively low accuracy. This research tests how combining the Naive Bayes classifier using Chi-Square as its feature selection, accompanied by Laplace Smoothing, may improve its accuracy. The tests classify 600 high school biology exam questions in Bahasa Indonesia into Bloom's Taxonomy of cognitive domain. Test results show an increase of 39.6% accuracy is produced from 21.03% to 60.63% when 100 new data introduced. On the other hand, a 22.19% increase is gained from 53.75% to 75.94% when a mix of testing data (10, 20, 50, 100, 250) is applied.**

Keywords—bloom's taxonomy, smoothing, naive bayes, chi-square, feature selection, examination.

I. Introduction

This research automatically classifies high school biology exam questions using the Naïve Bayer classifier. Naïve Bayer classifier can swiftly and effectively classify big data. It is also famous for an algorithm to withstand noise in data. However, Naïve Bayes has several weaknesses. These weaknesses produce a lower accuracy. This research tests the combination of the Naïve Bayes classifier, Chi-Square feature selection, and Laplace Smoothing to tackle those weaknesses.

Naïve Bayes utilizes the Bayes theorem to calculate future probability using past experiences. The main characteristic of the Naïve Bayes classifier is the strong assumption of independence of each feature. But sometimes, applying Naïve Bayes may bear misclassification for scarce training data, resulting in a 0 probability value (the zero problem) [1]. Previous researches [2], [3] discover that the smoothing technique can minimize this weakness. Naïve Bayes classifier performs better when correlated features are filtered. Correlated features voted twice lead to the overemphasis of the importance of the correlated features [4]. Frequency-based, mutual information or information gain feature selection can be used as feature selection to increase the Naïve Bayes classifier's accuracy.

Previous research applies the Naïve Bayes classifier to bank marketing data [5] and reviews of the online fashion company [6]. Most of them prove that the combination of feature selection and smoothing increases accuracy significantly [7]. Hence, this research combines Chi-Square and Laplace Smoothing on the Naïve Bayes classifier to tackle the zero problem that may arise when classifying high school biology exam questions. The composition of the under-standing level of each question expresses the quality of exam questions. Bloom's taxonomy of the cognitive domain is used commonly in the academic field. It divides question types into six stages of the cognitive domain as a basis for exam questions classification. An automated exam questions classification will efficiently ensure a more qualified exam.

II. Research Method

A. Dataset

Data are collected from high school biology exam questions in Bahasa Indonesia from various sources, i.e: student worksheets, daily exams, and biology modules. There are 2 (two) types of datasets: 600 exam questions that are divided equally into 6 classes, and random questions from the first dataset where 100 of them is made class-less. The later dataset is used as testing data.

B. Cognitive Domain Bloom's Taxonomy

The pyramid in Fig. 1 describes levels or classes on the cognitive domain in Bloom's Taxonomy. The three first levels (from the bottom) can be classified as Lower Order Thinking Skills, while the next three levels are Higher Order Thinking Skills. Both are categories used in the categorization process. Every class has a special feature on each level. Table I presents 6 classes that are used, and operational verbs corresponding to each class. This table is the reference for teachers in the manual process. To simplify the naming convention, every class is denoted as C1 to C6 [8].

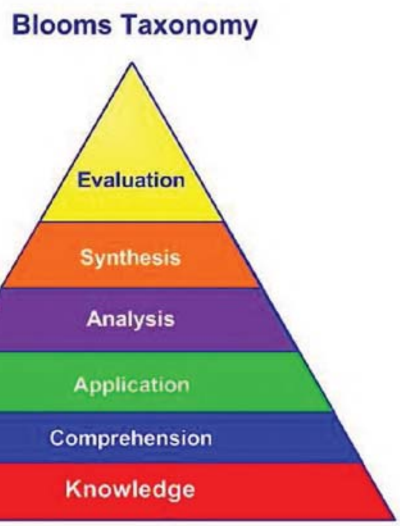

Fig. 1. Bloom's Taxonomy (Cognitive Domain)

TABLE I. CLASSES AND OPERATIONAL VERBS

Category	Example of operational verbs
C1	Mengutip, menyebutkan, mendaftar, menunujukkan, mengidentifikasi, melabeli, memasangkan, menamai, menandai, membaca, menyadari, menghafal, mencatat, mengulang memilih, menulis.
C2	Menerangkan, menjelaskan, menterjemahkan, menguraikan, mengartikan, menyatakan kembali, menafsirkan, menginterpretasikan, mendiskusikan, menyeleksi, mendeteksi, melaporkan, menduga, mengelompokkan, memberi
C3	Memilih, menerapkan, melaksanakan, mengubah, menggunakan, mendemonstrasikan, memodifikasi, menginterpretasikan, menunjukkan, membuktikan, menggambarkan, mengoperasikan, menjalankan, memprogramkan, mempraktekkan, memulai
C4	Mengkaji ulang, membedakan, membandingkan, mengkontraskan, memisahkan, menghubungkan, menunjukan hubungan antara variabel, memecah menjadi beberapa bagian, menyisihkan, menduga, mempertimbangkan
C5	Mengkaji ulang, mempertahankan, menyeleksi, mempertahankan, mengevaluasi, mendukung, menilai, menjustifikasi, mengecek, mengkritik, memprediksi, membenarkan, menyalahkan.
C6	Merakit, merancang, menemukan, menciptakan, memperoleh, mengembangkan,memformulasikan, membangun,membentuk, melengkapi,membuat, menyempurnakan, melakukan, inovasi, mendisain, menghasilkan karya.

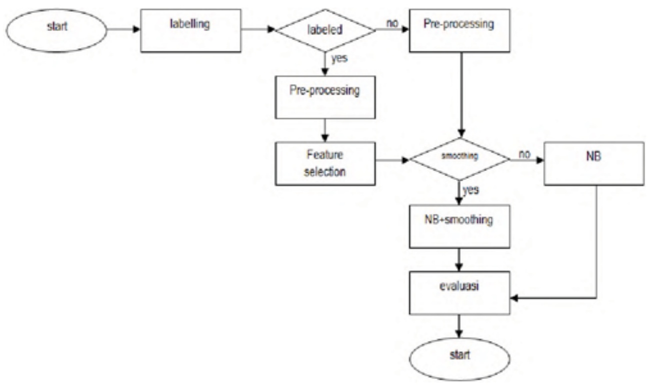

Fig. 2. Process Flow

C. Process Flow

This research uses labelled and unlabelled data. Labeled data is used as training data. As described in Fig. 2 we will analyze smoothing performance and test effects of the smoothing method to Naïve Bayes and Chi-Square feature selection. The feature selection method is applied to labelled data to dismiss irrelevant features.

The output of the Naïve Bayes classifier is a probability value that reflects the confidence of data to be labelled as a certain category. If variable d has to be classified, NBC (Naive Bayes Classifier) assumes that the features are independent, and variable c denotes the class, therefore $p(c_i)$ and $p(d|c_i)$ can be expressed as follows:

$$p(c_i) = \frac{|c_i|}{|C|} \quad (1)$$

$$p(d|c_i) = \prod_{k=1}^{|d|} p(w_k|c_i) \quad (2)$$

To improve the performance of the Naïve Bayes classifier, and to minimized misclassification, a smoothing method should be applied. This research uses Laplace Smoothing.

D. Laplace Smoothing

Laplace Smoothing is a common smoothing method, often called default smoothing, also the oldest smoothing method ever implemented on Naïve Bayes Classifie [9]. Laplace smoothing is also called Add-one smoothing because it adds 1 to every calculated token frequency. Naïve Bayes with Laplace Smoothing can be formulated as follows:

$$P(t|c) = \frac{T_{ct}+1}{\sum_{t' \in V} T_{ct'}+|V|} \quad (3)$$

Where $P(t|c)$ is a probability of a word appear in a certain class, and $|V|$ is unique words in all classes.

E. Chi-Square

Chi-Square or X^2 is a feature selection method that can be classified as a filter method approach [10]. Generally, Pada this feature selection method is used in preprocessing step.

$$X^2(t,c) = \frac{N(A\times D - B\times C)^2}{(A+B)\times(C+D)\times(A+c)\times(B+D)} \quad (4)$$

Variable t denotes a word being testing on class c, N denotes the count of training documents, A denotes documents in class c that contain word t, B denotes documents that are not in c but contain t, C denotes documents in c but has no t in it, then D denotes documents that are not in c and don't contain t [11].

F. Preprocessing

Broadly speaking, pre-processing step is conducted in an orderly fashion, as described in Fig. 3. The tokenizing step is run after the labelling process on data is done, continued by the case-folding step, stop word removal dan ended by the stemming step as the final result of the pre-processing flow. After pre-processing step is finished, the next steps are selecting features for the next step, deletion, and selection of features according to the threshold. Features that are not eligible will be deleted.

G. Evaluation

The evaluation step is an important step in research to quantify the performance of series of methods that are conducted. F-measure is used in this research. F-measure calculates Recall and Precision on every test. Recall is the number of relevant documents retrieved by a search divided by the total number of existing relevant documents, while precision is the number of relevant documents retrieved by a search divided by the total number of documents retrieved by that search.

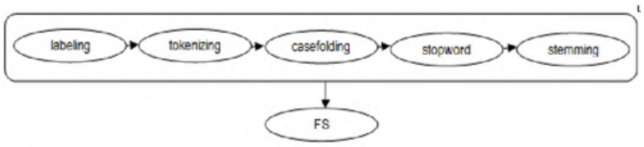

Fig. 3. Preprocessing

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad (5)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \quad (6)$$

$$\text{F} - \text{measure} = \frac{2\text{PR}}{\text{P+R}} \quad (7)$$

Parameter *TN*, *TP*, *FP*, *FN* are obtained from the Confusion Matrix in Table II. According to Han dan Kamber (2015), the Confusion Matrix can be used to analyse how well a classifier classifies different tuples.

III. Result and Evaluation

This research's goal is to analyze the performance of Laplace smothing on Naïve Bayes classifier using Chi-Square as a feature selection method to categorized exam questions into 6 (six) different classes based on cognitive domain level in Bloom's Taxonomy. There are 2 (two) datasets used. The first dataset contains labeled data, while the second dataset is a portion of labeled and unlabeled data from the first dataset. F-measure for every test is lined up in Table IV. The threshold for Chi-Square is 0,5. The test result of the first dataset Laplace smoothing manages to reach an F-measure of 60,363. This means an increase of 39,601 from Naïve Bayes without smoothing. Table III also shows that the size of training data is highly correlated to the F-measure value. Changes of F-measure using the first dataset can be seen in Fig. 4.

TABLE II. Confusion Matrix

		Prediction	
		Yes	*No*
Actual	Yes	TP	TN
	No	FP	FN

TABLE III. Test Result on Dataset 1

Training Set	Training Data Count	F-measure	
		NB-CHI	*NB-CHI-LAP*
10%	60	8,14	57,27
20%	120	19,61	64,09
30%	180	19,73	65,30
40%	240	22,91	63,21
50%	300	20,57	61,69
60%	360	18,99	57,51
70%	420	24,15	62,49
80%	480	26,60	63,18
90%	540	21,21	48,74
100%	600	28,44	62,88
Average		**21,035**	***60,636***

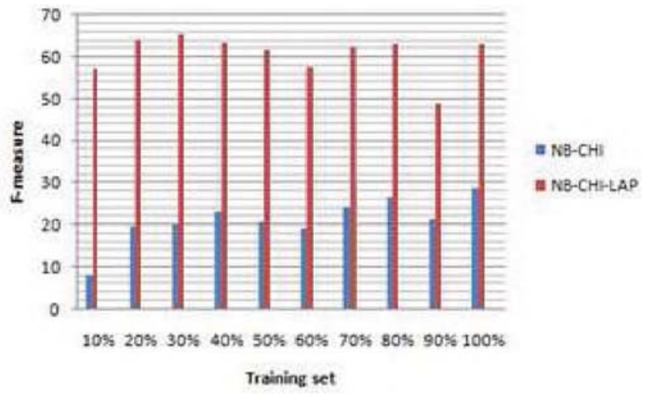

Fig. 4. Test Result of Dataset 1

TABLE IV. Test Result on Dataset 2

Training Set	F-measure	
	NB-CHI	*NB-CHI-LAP*
10	59,68	70,80
25	44,29	65,85
50	53,92	82,99
100	58,31	82,31
250	52,59	77,76
Average	53,758	75,942

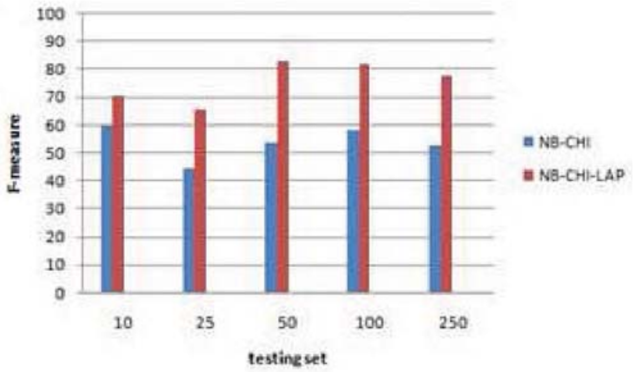

Fig. 5. Test Result of Dataset 2

The second test is run on the second dataset. Data being tested are 25, 50, 100, and 250 exam questions. This test tries to analyse how far smoothing can relabel data that are made unlabelled.The result of the first and second tests are likewise. The second test result indicates that Laplace Smoothing is proven to increase the performance of the Naïve Bayes classifier can be seen in Fig. 5.

Conclusion

The smoothing method successfully increases the performance of Naïve Bayesian Chi-Square. By using Laplace smoothing an increase of 39,601 is produced, compared to traditional Naïve Bayesian Chi-Square. This smoothing method is also proved to minimize misclassification errors.

References

[1] F. Indriani and D. T. Nugrahadi, "Comparison of Naive Bayes smoothing methods for Twitter sentiment analysis," in *2016 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, 2016, pp. 287–292.

[2] S. Aggarwal and D. Kaur, "Naive Bayes Classifier with Various Smoothing Techniques for Text Documents," *Int. J. Comput. Trends Technol.*, vol. 4, no. 4, pp. 873–876, 2013.

[3] Q. Yuan, G. Cong, and N. M. Thalmann, "Enhancing naive bayes with various smoothing methods for short text classification," in *Proceedings of the 21st International Conference on World Wide Web*, 2012, pp. 645–646.

[4] H. Li, "Machine learning for the subsurface characterization at core, well, and reservoir scales." 2020.

[5] S. Ernawati and E. R. Yulia, "Implementation of The Naïve Bayes Algorithm with Feature Selection using Genetic Algorithm for Sentiment Review Analysis of Fashion Online Companies," in *2018 6th International Conference on Cyber and IT Service Management (CITSM)*, 2018, pp. 1–5.

[6] T. Yang, K. Qian, D. C.-T. Lo, Y. Xie, Y. Shi, and L. Tao, "Improve the prediction accuracy of Naïve Bayes classifier with association rule mining," in *2016 IEEE 2nd International Conference on Big Data Security on Cloud (BigDataSecurity), IEEE International Conference on High Performance and Smart Computing (HPSC), and IEEE International Conference on Intelligent Data and Security (IDS)*, 2016, pp. 129–133.

[7] G. Qiang, "Research and improvement for feature selection on naive bayes text classifier," in *2010 2nd International Conference on Future Computer and Communication*, 2010, vol. 2, pp. V2-156.

[8] C. ann" Ratanamahatana and D. Gunopulos, "Feature selection for the naive bayesian classifier using decision trees," *Appl. Artif. Intell.*, vol. 17, no. 5–6, pp. 475–487, 2003.

[9] Z. H. Kilimci and M. C. Ganiz, "Evaluation of classification models for language processing," in *2015 International Symposium on Innovations in Intelligent SysTems and Applications (INISTA)*, 2015, pp. 1–8.

[10] Y. Yang and J. O. Pedersen, "A comparative study on feature selection in text categorization," in *Icml*, 1997, vol. 97, no. 412–420, pp. 35–44.

[11] D. A. Abduljabbar and N. Omar, "Exam questions classification based on Bloom's taxonomy cognitive level using classifiers combination," *J. Theor. Appl. Inf. Technol.*, vol. 78, no. 3, pp. 447–455, 2015.