Text Mining Approach Using TF-IDF and Naive Bayes for Classification of Exam Questions Based on Cognitive Level of Bloom's Taxonomy

Annisa Aninditya
Department of Information Systems,
Telkom University,
Bandung, West Java, Indonesia, 40257
annsdy@student.telkomuniversity.ac.id

Muhammad Azani Hasibuan
Department of Information Systems,
Telkom University,
Bandung, West Java, Indonesia, 40257
muhammadazani@telkomuniversity.ac.id

Edi Sutoyo
Department of Information Systems,
Telkom University,
Bandung, West Java, Indonesia, 40257
edisutoyo@telkomuniversity.ac.id

***Abstract*— Bloom's Taxonomy is a unity of three domains, which are divided into lower orders and high orders based on the Bloom Taxonomic Cognitive Domain, the level is used to classify learning objectives and serve as benchmarks for evaluating student achievement. Basically, an evaluation of student achievement can be done by giving questions on exam activities. The questions given are then classified according to the level in the Cognitive Domain. However, because the number of questions is too many and the classification is still manual, it causes the classification results are not accurate and inconsistent. Therefore, the employing of the Naive Bayes Classifier in classifying exam questions based on levels in the Cognitive Domain can be a solution. This study uses real-world dataset collected from mid-terms and final exams questions taken from Department of Information Systems, Telkom University from the academic year 2012/2013 to 2018/2019. In particular, we examined Words, Characters, and N-gram as indexing terms. The results showed that the classification using Naïve Bayes and TF-IDF with N-gram as indexing terms achieved precision of 85% and recall of 80%.**

Keywords—Bloom's Taxonomy, Exam Questions, Naive Bayes, Text Mining, TF-IDF, Classification

I. Introduction

There are many ways of assessment used to find out the development of learning. The written exam is the most commonly used assessment and has an important role in the effort to test the overall results of the cognitive level of students in each semester [1]. The most effective way of questions that have been described in [2] is a problem to help students achieve the desired learning outcomes. Examples of the arrangement of Bloom's Taxonomy which are widely used in education [3], [4] build questions, from computer science Domain itself, and Bloom's Taxonomy enhances curriculum design and assessment. The composition of Bloom's Taxonomy is known to be divided into three domains, namely: Cognitive, Affective, and Psychomotor Learning.

Generally, to identify the level of questions based on the level of Cognitive Level of Bloom's Taxonomy is done manually. The process of classifying questions manually requires considerable time for large data sizes. In addition, differences in perception in classifying the questions resulted in a classification process manually to be varied. As a solution to these problems, the automatism classification process can be done using Natural Language Processing [5], [6].

This study aims to determine the effectiveness in classifying questions based on the level of Cognitive Level of Bloom's Taxonomy using the Naive Bayes Classifier. In addition, as a lecturer solution in making questions that will be a benchmark for assessing students' understanding of the material given based on learning objectives. In this study, the dataset used uses questions from mid-term exams and final-exams given by lecturers which later will be classified manually and then entered into preprocessing such as techniques in tokenizing, stemming, filtering, and also feature extraction is made as a set of processes and then questions existing data will be converted into numerical data called vector features, after that it will be inputted into the Naive Bayes building model. The results of the study will be a model for classifying questions based on the level of the Cognitive Level of Bloom's Taxonomy.

The remainder of this paper is organized as follows. In section 2, the literature review is presented. The proposed approach is elaborated in Section 3. The result and discussion are discussed in Section 4. Finally, we conclude our works in Section 5.

II. Literature Review

A. Data Mining

Data mining is a process that uses statistical techniques, mathematics, artificial intelligence, machine learning to extract and identify useful information and related knowledge from various large databases [7]. Data mining has been widely applied in various fields such as education, engineering, economics and business, social, healthcare, and etc [8]–[17]. There are two models in data mining, namely supervised learning and unsupervised learning. Supervised learning is a model of data mining for a data set that has a label and is more widely used for predicting the output of new values entered into existing data sets. Examples of methods using this model are classification and regression. Instead, unsupervised learning is used to find a certain pattern that exists in a set of existing data sets [18].

B. Text Mining

Text mining can be broadly defined as a process of digging information where a user interacts with a group of documents using analytical tools which are components in data mining, one of which is categorization. Text mining is a classification method that is a variation of data mining trying to find interesting patterns from a large collection of textual data. Text mining [19]–[21] is grouped into 7, namely; (a) Information search and acquisition, (b) Document classification, (c) Document classification, (d) Web mining, (e) Information extraction, (f) Natural language processing, and (g) Concept extraction. In stages, Text mining on

978-1-7281-2516-9/19/$31.00 ©2019 IEEE

documents or a text is carried out as follows: (a) The tokenizing stage, (b) The Filtering Stage, and (c) The Stemming Stage.

C. *Natural Language Processing*

Natural Language Processing (NLP) is a process of language identification that is done to facilitate humans in communicating with computers using human language [22], [23]. To be able to understand natural languages it is very important to be able to know and distinguish the stages in taking or extracting understanding from the text [24].

The system of natural language processing involves at least one or all of the following levels of analysis, namely;

- Phonetic/phonological level related to pronunciation
- Morphological level relating to the smallest element of the word that contains its own meaning and the prefix and suffix
- Syntactic level related to sentence structure
- Semantic levels that relate to the meaning of words and sentences
- Discourse level related to the structure of words using the structure of the document
- Pragmatic level relating to information that comes from outside the object under study, for example data from outside the document.

D. *Bloom's Taxonomy*

Bloom's taxonomy was introduced by Benjamin Bloom, Englehart, Furst, Hill and Krathwohl [25] in the 1950s and then revised by Krathwohl [4]. Bloom's Taxonomy discusses the setting of learning objectives in the educational process designated by educators for students. Bloom's taxonomy makes teachers more aware of the content and processes taught and values and shows the difference between what is taught and what is valued. Bloom's Taxonomy can serve as a guide for developing and expanding learning and assessment activities by supplying concrete awareness about the content and teaching process in defining what is essential in the development of learners [26]. Bloom's taxonomy is used to identify a person's ability from the lowest level to the highest level. This concept is divided into three domains of intellectual ability, namely cognitive, affective, and psychomotor.

E. *Cognitive Learning*

Higher-Order Thinking originates from the Cognitive Level of Bloom's Taxonomy which was introduced in 1956. The Cognitive Domain involves the knowledge and development of intellectual skills which includes the recall or recognition of specific facts, procedural patterns, and concepts that serve to develop intellectual abilities and skills [27], [28]. There are six levels in the taxonomy from the lowest to the highest order process, the first three levels in the Cognitive process, which deal with low-level thinking skills that are important in laying the foundation for deeper understanding or also called Lower Order. The last three use higher-order thinking skills, or also called Higher Order, (a) knowledge or ability to know or recall information; (b) comparison or ability to explain understanding by describing paraphrase, etc.: (c) application of learning information to solve a problem or answer a question; (d) analyze or explain the problem into parts; (e) evaluating or evaluating an idea using clear criteria; and (f) reconstructing knowledge into new patterns. Fig 1 shows the level of Cognitive Domain of Bloom's Taxonomy, which based on 3 levels from the low is a Low Order, and the next 3 levels are a High Order and can be explained as follows.

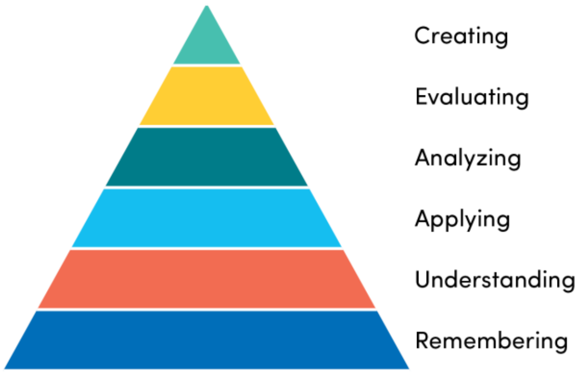

Fig. 1. The Cognitive Domain of Bloom's Taxonomy [29]

III. The Proposed Approach

In this section, a brief explanation of the proposed method and dataset are presented. This approach is comprised of three main steps: text preprocessing (labeling the dataset, tokenizing, stemming, and filtering), feature selection based on TF-IDF, and text classification using Naive Bayes Classifier. Then the proposed approach for classifying exam questions based on the level of Cognitive Domain of Bloom's Taxonomy is investigated. Fig 3 shows the proposed approach used in this study.

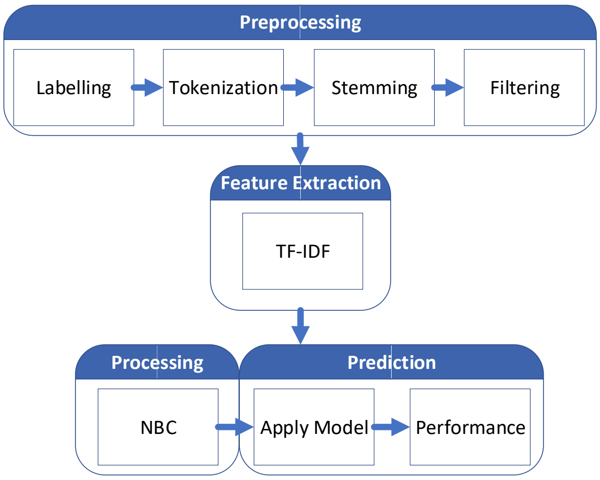

Fig. 2. The proposed approach

A. *Text Preprocessing*

Preprocessing is the first step in classification. The purpose of preprocessing is to interpret sentences or documents into feature vectors by turning text into words [30]. A text cannot be processed directly by a search algorithm, therefore it requires preprocessing text to convert text to numeric data. The preprocessing method plays a big role in the text mining process because it is at this stage that will determine the results of the data to be processed in the classification process. There are several stages in conducting the preprocessing process, namely:

- Tokenization

In the tokenizing process, the decomposition of words is meant to divide the text into collections of words without regard to the relationship between one word with another word, and the role and position of the sentence, which is the process of separating words in a text.

- Stemming

Stemming is the process of removing affixes, prefixes, suffixes, and prepositions so that the term can be the basic form of words. With the discovery of the basic word from the term, it can also be found the intensity of the appearance of the term in each document through the indexing process. For example, the word used, uses, is useful, will be stemmed into its root word "use". The stemming algorithm for one language is different from the stemming algorithm for another language. The stemming algorithm for Indonesian text is more complicated/complex when compared to stemming in English text because there are variations in the affix that must be removed to get the root word of a word.

- Filtering

Filtering is a Stop word Removal method, which is the process of using tokenizing data. This stage will be done by removing the words for the classification process such as the words: di, which, and others that are used as a limit on the size of the words that will be done next process. This limitation starts from the minimum and maximum size. Cleansing words here such as pronouns, prepositions, conjunctions. Stop words removal is the process of eliminating unimportant words in the description through checking the results of the description parsing words whether included in the list of words is not important (stop list) or not Lower case.

- Feature Extraction

Feature Extraction is the process of writing a word list from text data and then turning it into a set of features used for classification. The use of Feature Extraction as a measurement of the observed process, the method used in this process also has developments. In this case, the method used as Feature Extraction is Term Weighting. Term Weighting is an advanced process by calculating the weight of a term that is in the question data. One of the Term Weighting techniques used for classification is the Term Frequency - Inverse Document Frequency (TF-IDF) technique. The application of TF-IDF is based on research [31] which shows that the application of TF-IDF used for word weighting is very influential in classification and getting better and optimal results.

B. Term Frequency – Inverse Document Frequency (TF-IDF)

TF-IDF is a method that is integration between term frequency (TF), and inverse document frequency (IDF). The function of the TF-IDF method is to look for a representation of the value of each document from a training data set wherein later a vector is formed between the document with the word (documents with terms) and then for similarity between the document and the cluster will be determined by a prototype vectors, also called centroid clusters. Term Frequency is calculated using Equation with the $i-th$ term frequency is the frequency of the $i-th$ term appearing in the *j*-th document. Inverse Document Frequency (IDF) is the logarithm of the ratio of the total number of documents in the corpus to the number of documents that have the term in question as written mathematically in the equation. The following formula is used to assign values to words using TF-IDF:

- Term Frequency (TF)

Term Frequency is the simplest way of weighting terms (words). The weight of the word t in the document is given using the following equation

$$w_{ij} = tf_{ij} \; x \; idf \tag{1}$$

- Inverse Document Frequency (IDF)

Inverse Document Frequency is a calculation of the appearance of words in a collection of documents. The IDF factor of a *t* word is given using the following equation.

$$idf = log \frac{N}{df_j} \tag{2}$$

- Words

Words is an extracted and selected word that will be used as a term in TF-IDF. The term used will be calculated based on the TF and IDF, after that the results of the term weighting assessment will be used as a feature in the prediction model that will be made at the classification stage.

- N-Gram

N-Gram is a series of N words from the order of words extracted and selected, with n represented starting from 1. When n is 1, it shows unigram; when n is 2, it shows bigram; and when n is 3, it shows the trigram. In this study using n is 2, which shows bigram. The bigram word series will later be used as terms in TF-IDF [32].

C. Naive Bayes Classifier (NBC)

Naive Bayes is a simple probabilistic classification that calculates a set of probabilities by adding up the frequency and combination of values from a given dataset. The algorithm uses the Bayes theorem and assumes all independent or non-interdependent attributes given by values to class variables. The main characteristic of the Naive Bayes Classifier (NBC) is a very strong assumption of the independence of each condition, where each attribute that builds each class is mutually exclusive. The value of these NBC features can be in the form of static data or in the form of categories. so that in the process already obtained definite results as well based on the theory from Bayes as in Equation 3 below,

$$P(H|X) = \frac{P(X|H) \times P(H)}{P(X)} \tag{3}$$

In the Naive Bayes method, the classification process requires some guidance to determine which class is suitable for the sample being analyzed. Therefore, the Naive Bayes method can be adjusted as shown in the Equation 4 as follows.

$$\mathrm{P(C|F1 \ldots Fn)} = \frac{\mathrm{P(C)P(F1 \ldots Fn|C)}}{\mathrm{P(F1 \ldots Fn)}} \tag{4}$$

D. Multinomial Naive Bayes

Multinomial Naive Bayes is one of the development models of Bayes algorithm which is influenced by a series of terms. This model is used for classifying text or documents. The calculation uses the number of terms in the document, the opportunities that arise between the terms are independent. In the Multinomial Naive Bayes Classifier formula, the class of documents is determined by the words that appear and the appearance of the words of each document, as shown in Equation 5 as follows.

$$P(c|d) = \frac{P(d|c)P(c)}{P(d)} \tag{5}$$

Taking into account the term opportunity value in documents in class c can be calculated by Equation 6 below,

$$P(c|d) = P(c)\prod P(fk|c)\ 1 \le k \le nd \tag{6}$$

To calculate parameters $P(fk|c)$, will be used the Equation 7 below.

$$P(fk|c) = \frac{count(fk,c)+1}{count(c)+|v|} \tag{7}$$

Whereas, to calculate parameters $P(c)$, will be used the Equation 8 below.

$$P(c) = \frac{N_c}{N} \tag{8}$$

E. Multi-Variate Bernoulli

Multivariate Bernoulli is a Naive Bayes Classifier function that uses binary elements to take the value of 1 if the corresponding word is in the document and 0 if the word does not exist.

$$P(fk|c) = \frac{Tct+1}{Tc+\Sigma c} \tag{9}$$

Meanwhile, to calculate the probability value of a sentence to a class can be calculated with the equation described in II.14,

$$P(c|d) = P(c)\Pi_{i=1}^{N}P(fk|c) \times \Pi_{i=1}^{M}(1-\mathrm{P(fk'|c)}) \tag{10}$$

F. Classifier Performance

Evaluating performance is not only to choose the best classification but also to verify that the classification is designed to meet the design requirements and to identify the need to improve the classification component. One of the metrics commonly used to evaluate classification performance is evaluation metrics, namely Precision and Recall, which are calculated using the Confusion Matrix.

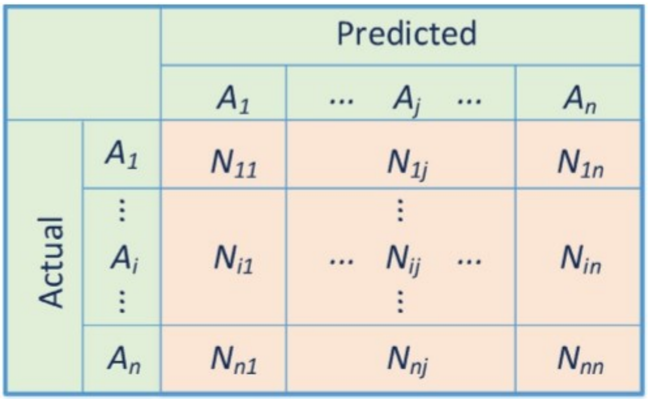

Fig. 3. Confusion Matrix [33].

$$precision = \frac{number\ of\ correct\ positive\ prediction}{number\ of\ positive\ prediction}$$

$$Recall = \frac{number\ of\ correct\ positive\ prediction}{number\ of\ positive\ examples}$$

IV. Result and Discussion

A. Dataset

In this study, the dataset used is a collection of exam questions from the mid-term and final-exams from Department of Information System, Telkom University academic year from 2012/2013 to 2018/2019. The data will then be classified manually and then entered into preprocessing such as techniques in tokenizing, stemming, filtering, and also feature extraction is made as a set of processes and the questions will be converted into numerical data called vector features, after that, it will be inputted into the Naive Bayes building model. The results of the study will be an approach for classifying exam questions based on the level of the Cognitive Domain of Bloom's Taxonomy.

After the question data has been collected then determine the label of each question. Labeling based on the level of Cognitive Level of Bloom's Taxonomy which is divided into two namely, Lower Order (LO) and High Order (HO). Table I is an example of the results of the labeling that has been done by 5 Annotators (AN1-AN5).

TABLE I. Example of The Process of Labeling Data by 5 Annotators

No	Question	AN1	AN2	AN3	AN4	AN5
1	Q 1	HO	LO	LO	LO	HO
2	Q 2	LO	LO	LO	LO	LO
3	Q 3	LO	LO	HO	LO	LO

After the labeling by the Annotator is completed then data is collected and the label taken is the most of the 5 labels obtained, can be seen in Table II which is an example of questions and labels used as a dataset.

TABLE II. Example of Dataset

No	Question	Label
1	Question 1	LO
2	Question 2	LO
3	Question 3	LO

After completing the labeling the data, the comparison the number of the High Order and Low Order labels is shown in Fig 4 with the High Order comparison of 143 questions and Low Order of 157 questions.

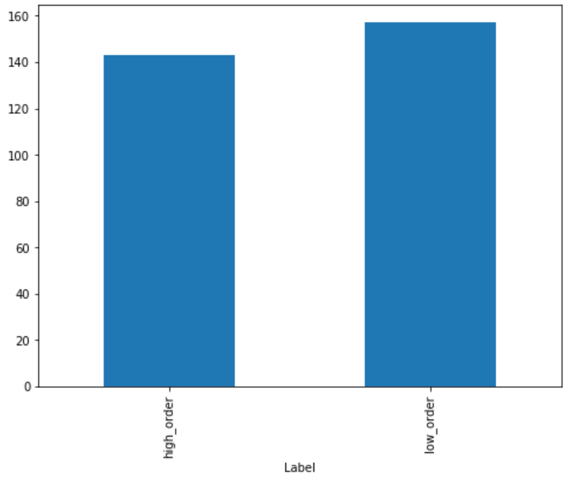

Fig. 4. Comparison of the number of data that has been successfully labeled

B. Performance Measure

In the splitting corpus step, the corpus (processed data) will be divided into two parts, Data Training, and Data Testing. Data Training is used as a classification model and the results of classifying Data Training will be used as a model to predict labels on Data Testing. Data is divided into 80% Data Training and 20% Data Testing. Distribution of data using Cross-Validation Score. in this study the dataset was participated in into 10 sections, which will be used for classification.

In this study, the NBC Method will be divided into 2 stages, namely the Training Stage and Testing Phase. At the training stage, using Data Training which will be carried out an analysis of the sample document in the form of vocabulary selection, which is a word that might appear in the sample collection, which can later be a representation of questions for each label. After doing the testing phase, at this stage using Data Testing that will determine the label of the question based on vocabulary that already represents the type of question for each label. In this study using k = 10 with the TF-IDF feature which will produce an average value of accuracy. Table III is the result of the 10 Fold-Cross Validation used at the Words TF-IDF feature which yields an average accuracy of 78%.

TABLE III. THE RESULTS OF 10 FOLD-CROSS VALIDATION USED ON THE WORDS TF-IDF

No.	*K*	Accuracy
1	1	68%
2	2	58%
3	3	87%
4	4	83%
5	5	83%
6	6	67%
7	7	80%
8	8	86%
9	9	79%
10	10	90%
Average		78%

Table IV shows the results of the 10 Fold-Cross Validation using the N-Gram TF-IDF feature which produces the lowest average accuracy of some TF-IDF level features because it only produces 74%.

TABLE IV. THE RESULTS OF 10 FOLD-CROSS VALIDATION USED ON THE N-GRAM TF-IDF

No.	*K*	Accuracy
1	1	58%
2	2	65%
3	3	84%
4	4	83%
5	5	73%
6	6	77%
7	7	73%
8	8	76%
9	9	76%
10	10	79%
Average		74%

Table V shows the results of the 10 Fold-Cross Validation by using the Characters TF-IDF feature, after classifying each fold it will get an average accuracy rate. On the Characters Level feature, the average accuracy rate only achieves 76%.

TABLE V. THE RESULTS OF 10 FOLD-CROSS VALIDATION USED ON CHARACTERS TF-IDF

No.	*K*	Accuracy
1	1	58%
2	2	68%
3	3	90%
4	4	80%
5	5	80%
6	6	77%
7	7	77%
8	8	76%
9	9	83%
10	10	76%
Average		76%

After getting an average accuracy value based on the 10 Fold-Cross Validation, to see the effectiveness of the classifier performance an accuracy calculation based on precision and recall is calculated based on the Confusion Matrix. As shown in Fig 5, the classification result display is based on the Words TF-IDF, N-Gram TF-IDF, Characters TF-IDF, respectively.

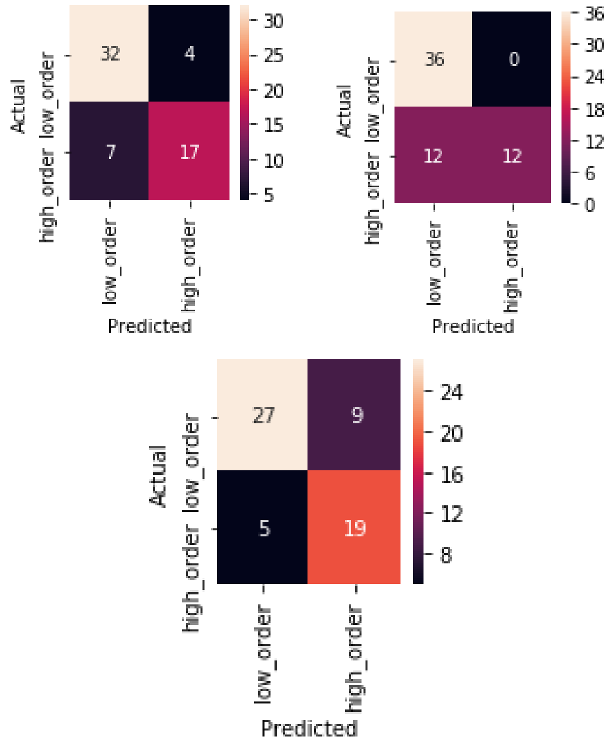

Fig. 5. Confusion Matrix based on Words TF-IDF, N-Grams *TF-IDF,* Characters *TF-IDF, respectively*

In the testing phase, the Naive Bayes Classifier shows conformity based on the results of the classification using the N-Gram TF-IDF feature, it can be analyzed that the use of the TF-IDF feature in making a fictional model greatly influences the results. Based on the results of the classification shows the performance of the prediction model is also influenced by the type of questions used. The results of the tests shown in Tables VI can be seen that in developing a classification of questions based on the system in Bloom's Taxonomy it is recommended to pay more attention to the questions used, which pay attention to terms or keywords from the level of Cognitive Level of Bloom's Taxonomy in question selection.

TABLE VI. NAIVE BAYES CLASSIFIER PERFORMANCE RESULTS

Approach	Precision	Recall
Naive Bayes Classifier Performance results using Words TF-IDF	82%	82%
Naive Bayes Classifier Performance results using Characters TF-IDF	78%	77%
Naive Bayes Performance Classifier results using N-Gram TF-IDF	85%	80%

Based on the results of the test, it can be seen the suitability of Naive Bayes Classification as a model of prediction of questions and being an optimal classifier because from the results obtained, when using Words TF-IDF the precision is 82%, the recall is 82%. While the result obtained for N-Gram TF-IDF, the precision is 85% and the recall is 80%, and the result obtained for Characters TF-IDF, the precision is 78% and the recall is 77%, respectively.

V. Conclusion

Based on the results of this study, Naive Bayes Classifier can be effectively used as a prediction model to classify the types of questions based on high and low orders of Bloom's Taxonomy Cognitive Domain. Analysis of the application of Naive Bayes Classifier by using features for different Term Weighting can affect the results of the classification with the results of different levels of accuracy. Based on the results of this study it can be concluded that Naive Bayes using the feature of TF-IDF N-Gram Level is effective so that it can achieve the highest accuracy precision of 85%.

References

[1] N. Omar *et al.*, "Automated Analysis of Exam Questions According to Bloom's Taxonomy," *Procedia - Soc. Behav. Sci.*, vol. 59, pp. 297–303, Oct. 2012.

[2] A. J. Swart, "Evaluation of Final Examination Papers in Engineering: A Case Study Using Bloom's Taxonomy," *IEEE Trans. Educ.*, vol. 53, no. 2, pp. 257–264, May 2010.

[3] E. J. Furst, "Bloom's Taxonomy of Educational Objectives for the Cognitive Domain: Philosophical and Educational Issues," *Rev. Educ. Res.*, vol. 51, no. 4, pp. 441–453, Dec. 1981.

[4] D. R. Krathwohl, "A Revision of Bloom's Taxonomy: An Overview," *Theory Pract.*, vol. 41, no. 4, pp. 212–218, Nov. 2002.

[5] K. Mahesh and S. Nirenburg, "Semantic classification for practical natural language processing," 1995.

[6] R. Kuhn and R. De Mori, "The application of semantic classification trees to natural language understanding," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 17, no. 5, pp. 449–460, May 1995.

[7] J. Han, J. Pei, and M. Kamber, *Data mining: concepts and techniques*. Elsevier, 2011.

[8] I. Kaastra and M. Boyd, "Designing a neural network for forecasting financial and economic time series," vol. 10, no. 3, pp. 215–236, 1996.

[9] R. L. Grossman, C. Kamath, P. Kegelmeyer, V. Kumar, and R. Namburu, *Data mining for scientific and engineering applications*, vol. 2. Springer Science & Business Media, 2013.

[10] P. Giudici, *Applied data mining: statistical methods for business and industry*. John Wiley & Sons, 2005.

[11] D. Braha, *Data mining for design and manufacturing: methods and applications*, vol. 3. Springer Science & Business Media, 2013.

[12] H. Chiroma *et al.*, *An intelligent modeling of oil consumption*, vol. 320. 2015.

[13] E. Sutoyo, I. T. R. Yanto, Y. Saadi, H. Chiroma, S. Hamid, and T. Herawan, "A Framework for Clustering of Web Users Transaction Based on Soft Set Theory," in *Springer*, 2019, pp. 307–314.

[14] E. Sutoyo, R. R. Saedudin, I. T. R. Yanto, and A. Apriani, "Application of adaptive neuro-fuzzy inference system and chicken swarm optimization for classifying river water quality," in *Electrical, Electronics and Information Engineering (ICEEIE), 2017 5th International Conference on*, 2017, pp. 118–122.

[15] R. R. Saedudin, E. Sutoyo, S. Kasim, H. Mahdin, and I. T. R. Yanto, "Attribute selection on student performance dataset using maximum dependency attribute," in *Electrical, Electronics and Information Engineering (ICEEIE), 2017 5th International Conference on*, 2017, pp. 176–179.

[16] E. Sutoyo, I. T. R. Yanto, R. R. Saedudin, and T. Herawan, "A soft set-based co-occurrence for clustering web user transactions," *Telkomnika (Telecommunication Comput. Electron. Control.*, vol. 15, no. 3, 2017.

[17] R. R. Saedudin, H. Mahdin, S. Kasim, E. Sutoyo, I. T. R. Yanto, and R. Hassan, *A relative tolerance relation of rough set for incomplete information systems*, vol. 700. 2018.

[18] V. Kotu and B. Deshpande, "Predictive analytics and data mining: concepts and practice with rapidminer," 2014.

[19] M. Hearst, "What is text mining," *SIMS, UC Berkeley*, 2003.

[20] C. C. Aggarwal and C. Zhai, *Mining text data*. Springer Science & Business Media, 2012.

[21] K. B. Cohen and L. Hunter, "Getting started in text mining," *PLoS Comput. Biol.*, vol. 4, no. 1, p. e20, 2008.

[22] C. D. Manning, C. D. Manning, and H. Schütze, *Foundations of statistical natural language processing*. MIT press, 1999.

[23] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa, "Natural language processing (almost) from scratch," *J. Mach. Learn. Res.*, vol. 12, no. Aug, pp. 2493–2537, 2011.

[24] S. Feldman, "NLP meets the Jabberwocky: Natural language processing in information retrieval," *Online-west. THEN WILTON-*, vol. 23, pp. 62–73, 1999.

[25] B. S. Bloom and others, "Taxonomy of educational objectives. Vol. 1: Cognitive domain," *New York McKay*, pp. 20–24, 1956.

[26] D. Köksal and Ö. G. Ulum, "Language assessment through Bloom's Taxonomy," *J. Lang. Linguist. Stud.*, vol. 14, no. 2, pp. 76–88, 2018.

[27] J. P. Byrnes, *Cognitive development and learning in instructional contexts*. Allyn and Bacon Boston, 1996.

[28] A. G. Greenwald, "Cognitive learning, cognitive response to persuasion, and attitude change," *Psychol. Found. attitudes*, pp. 147–170, 1968.

[29] M. Forehand, "Bloom's taxonomy," *Emerg. Perspect. Learn. teaching, Technol.*, vol. 41, no. 4, pp. 47–56, 2010.

[30] A. Kao and S. R. Poteet, *Natural language processing and text mining*. Springer Science & Business Media, 2007.

[31] A. M. Kibriya, E. Frank, B. Pfahringer, and G. Holmes, "Multinomial naive bayes for text categorization revisited," in *Australasian Joint Conference on Artificial Intelligence*, 2004, pp. 488–499.

[32] J. Acs, L. Grad-Gyenge, and T. B. R. de Rezende Oliveira, "A two-level classifier for discriminating similar languages," in *Proceedings of the Joint Workshop on Language Technology for Closely Related Languages, Varieties and Dialects*, 2015, pp. 73–77.

[33] X. Deng, Q. Liu, Y. Deng, and S. Mahadevan, "An improved method to construct basic probability assignment based on the confusion matrix for classification problem," *Inf. Sci. (Ny).*, vol. 340–341, pp. 250–261, 2016.

978-1-7281-2516-9/19/$31.00 ©2019 IEEE