Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 3, No. 4, August 2022, Page 322−329
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v3i4.2079

Question Answering System at the Kingdom of Sumedang Larang with Naïve Bayes Method

Richo Fedhia Saldhi, Z K A Baizal*, Ramanti Dharayani

Informatics, School of Computing, Telkom University, Bandung, Indonesia
Email: [1]richofedhias@student.telkomuniversity.ac.id, [2,*]baizal@telkomuniversity.ac.id, [3]dharyani@telkomuniversity.ac.id
Email Penulis Korespondensi: baizal@telkomuniversity.ac.id
Submitted: **10/08/2022**; Accepted: **23/08/2022**; Published: **30/08/2022**

Abstract−The Sumedang Larang Kingdom is one of the kingdoms in Indonesia which was founded by Prabu Tajimalela in 721 AD. The Sumedang Larang Kingdom is known as the national history of Indonesia. Still, most of the current generation does not know the history of the Sumedang Larang Kingdom, especially the younger generation. Therefore, we developed a question-and-answer system to seek information about the Sumedang Larang Kingdom. With the development of information technology, research on question answering systems is applied to research on Biomedical Questions to produce correct answers. Our system will help literacy about the Sumedang Larang Kingdom for the younger generation, especially students, and increase Indonesian cultural assets. The QA system aims to generate and provide precise short answers to user questions by automatically using information extraction and natural language processing methods. To collect and create questions, we use the concept of ontology. In addition, we use the Natural Language Naïve Bayes method to answer user questions. We built a QA system that can help students find information about the history of the Sumedang Larang Kingdom. Based on the accuracy of the results of testing the method we propose. In our evaluation, we involve the Decision Tree method as the base model. We note that the accuracy of the Naïve Bayes method is higher than that of the Decision Tree. The accuracy result of Naïve Bayes at the ratio of 8:2 and 7:3 is 67%, while the Decision Tree is only 56%.

Keywords: Question Answering; Naïve Bayes; Ontology

1. INTRODUCTION

Historically, several kingdoms have been in Indonesia, including the Sumedang Larang Kingdom. The Sumedang Larang Kingdom was a kingdom founded in 721M by King Tajimalela. This kingdom is famous in Indonesian national history, especially in Tanah Parahiyangan, after the Pajajaran Kingdom ended. The kingdom of Sumedang Larang is widely praised, especially among the poets and poets who tell the story of Sumedang in fairy tales, rhymes, and even legends. However, now, many young people are not familiar with the history of the Sumedang Larang Kingdom. To make it easier for the younger generation to know the history of the Sumedang Larang Kingdom and with the development of information technology, we created a Question Answering System to make it easier for the public. In education, this system helps facilitate the search for information for students about the history of the Sumedang Larang Kingdom. Especially the younger generation, to find information about the history of the Sumedang Larang Kingdom. In addition, in the economic field, this system can increase the cultural assets of Indonesia. The QA system aims to generate and provide precise short answers to user questions by automatically analysing thousands of articles using information extraction and natural language processing methods. In previous research, the question answering system has implemented on. biomedical questions where the test was carried out with yes or no questions so that questions from users were limited [1]. Based on the research, we conducted research on the application of the Question Answering System for Information about the Sumedang Larang Kingdom. In this study, the use of yes or no questions was not the result of collecting user question data. While in our study, we use more flexible questions with the question dataset we have created to reach questions from users. To collect and generate questions, we use the concept of ontology. Ontology is part of information technology used for one domain so that it can generate questions [2]. Research in ontology-based questions has developed in various domains, such as music, biology and knowledge representation[3], [4]. Ontology presents information facts in their logical relationship, thus providing opportunities for the development of historical research [5]. We built a QA system based on a chatbot that can help students find information about the history of Sumedang Larang. After collecting and generating questions, there is a prediction process based on the respective questions and answers. in the process of searching for answers, we use the Naive Bayes method. We use the Naive Bayes method because in our research we classify several question sentence texts. This is based on several studies that apply the Naive Bayes method to classify text such as text-based chatbot applications that apply the Naive Bayes algorithm for the classification method.[6] Naive Bayes is a classification method rooted in Bayes' theorem where two separate events will later be classified. In testing, we compare the accuracy value between the method we use, namely Naive Bayes and the Decision Tree method. The Decision Tree method uses entropy to classify questions. Meanwhile, the Naive Bayes method uses multinomials to classify questions.

2. RESEARCH METHOD

Question Answering can identify accurate and concise answers to user questions. The test uses yes or no questions, using Meta thesaurus and frequency metrics. UMLS is used to extract exact answers from factoid questions [1].

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 3, No. 4, August 2022, Page 322−329
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v3i4.2079

The ontology used in the learning process will contribute to creating interactive and cognitive software solutions. Chatbots interact with users using pattern-based matching rules. Topics focused on the interactive use of methods in the teaching process and the development of methodological models for classroom work [[7]. The Ontology is also for increasing vocabulary so that it is easy to read by machines. Tested by calculating the score to determine the semantic relationship [8]. Ontology is also for creating an information retrieval system with the physics domain. Our tests calculate accuracy using precision values and recall [7]. The research about analyzing the comparison between the Naive Bayes method in the process of searching for answers the smoothness of financing concluded that in Naive Bayes, an accuracy rate of 50%.

2.1 Kingdom of Sumedang Larang

The Sumedang Larang Kingdom was increasingly recognized on the stage of Indonesian national history, especially in Tanah Parahiyangan, after the end of the Pajajaran Kingdom or the Sunda Kingdom in 1580. Sumedang was once revered, admired, and respected, not to mention the poets and poets telling Sumedang stories in their works, whether in titles, fairy tales, chronicles, rhymes, sagas, and even legends.

According to Carita Parahiyangan, the Tembong Agung Kingdom was led by Prabu Guru Haji Aji Putih, as the first king. It is said that Prabu Guru was still related to kinship with Prabu Sri Baduga Maharaja of the Sunda Kingdom at that time. He also had a son named Batara Tuntang Buana who later founded the Sumedang Larang Kingdom. When leading his new kingdom, Batara changed his name to Prabu Resi Tajimalela. Then he said, "Ingsun Medal Ingsun Madangan" (I was born I gave enlightenment), which is now the name SUMEDANG.

2.2 Question Answering (QA)

QA system is a system for enabling computers understand the questions asked by using natural language other than that QA system can respond to questions based on available information [9]. QA systems have special significance and advantages over machines search [10]. In QA, users simply type a question and display the list of web pages from the most suitable [11]. Users and analysts want targeted answers to their questions in the absence of addition of unrelated information [11]. So, the answer you get closely related to the question and in accordance with the expected answer desired.

In this study, the QA system was built using several stages, such as Text Preprocessing, Determination of similarity, and Answer Processing. Text Preprocessing serves to change the text into a term index, in which the question dataset that has been created will go through the preprocessing stage, i.e., stemming and stopword. Determination. of similarity is used to find the similarity of user questions with the method model that already contains a dataset of questions, subjects, objects, and keywords that will be matched with the dataset of answers that have been provided after the SPARQL query is processed. Answer processing functions retrieve answers from the ontology and give them to the user.

2.2 Text Preprocessing

Preprocessing stage is the stage to convert text into structured form. At this stage, stop word is performed to remove stop word in a text and stemming is to map or restore words to their basic form [12]. Stop words are words that are not processed into the algorithm such as which, to, in etc. The first stage of text preprocessing receives input from the user which was previously converted to lowercase. Lowercase is used to convert uppercase letters in sentences to lowercase. Furthermore, the results of the lowercase will be processed at the stopword stage which serves to delete words that often appear. After going through the stopword, the next step is stemming which serves to remove affixes in a sentence.

2.3 TF-IDF

TF-IDF is a statistical measurement that describes a word for a document so that it allows the word or term to replace the corpus in a document. TFIDF measurement can be done using the formula below.

$$W_{t,d} = tf_{t,d}.\log\frac{N}{df_t} \quad (1)$$

In the TFIDF formula, where $\boldsymbol{tf_{t,d}}$ is the number of occurrences of the term t in document d. $\boldsymbol{tf_{t,d}}$ measures the amount of information on a word or term in all documents and is calculated using the logarithm value of the number of documents in a corpus $\boldsymbol{N}$ and divided by the number of all documents containing $\boldsymbol{t}$. The more a term appears, the greater the $\boldsymbol{W}$ value will be.

2.4 Ontology

Ontology is a philosophy that was initially described as the science of being. However, in the language of modern computer science, ontology is defined as the essence of the concept and its related knowledge of ontology which is very useful for question classification systems, question analyzers, and answer extractors. Ontology is suitable for this QA system because ontology is more flexible in accepting questions from users. Several tools can retrieve

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 3, No. 4, August 2022, Page 322−329
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v3i4.2079

relevant information on ontologies, such as Protege. In this study, we use the Protégé tool because it is easier to use and user-friendly. Protege has several steps that must be follow by:

1. Create a new instance or class
2. Create the required properties
3. Add individual classes in each class or instance
4. Link existing properties into individual classes
5. Use SPARQL query to retrieve the required individual classes[13]

3. RESULT AND DISCUSSIONS

The method used by us in this study has several processes consisting of Preprocessing, Process Predict Questions, Query SPARQL and Ontology. In Figure 1 is an architecture system.

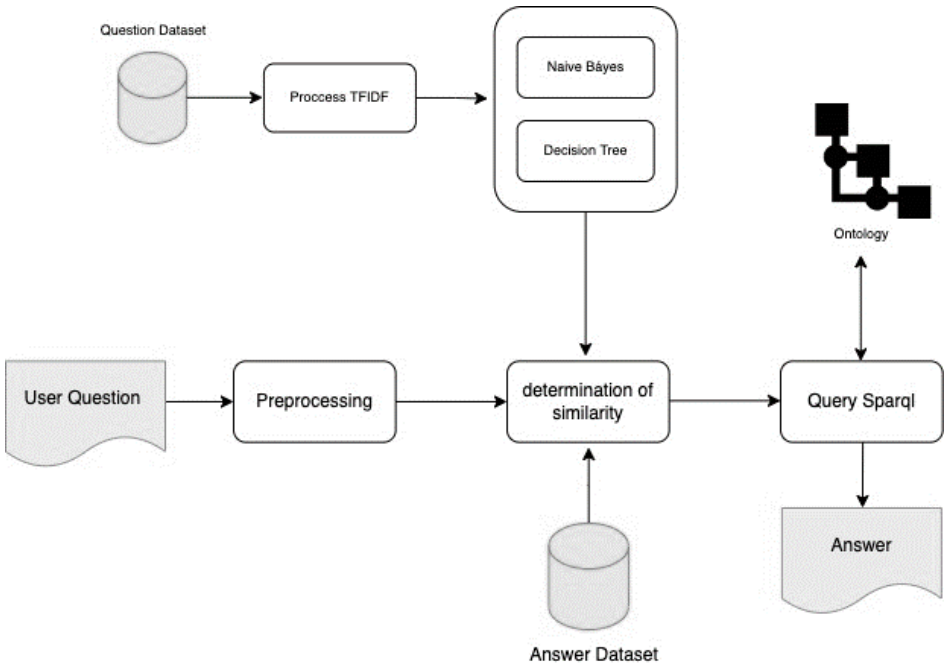

Figure 1. Architecture System

In Figure 1, the first question dataset is in the split ratio process to 8:2 and 7:3. After that, it enters the TF-IDF process, where the dataset that has been split based on a predetermined ratio will convert into a vector. After becoming a vector form, the following process is to enter the Decision method Tree using entropy and Naive Bayes using Multinomial. Questions from users will go through preprocessing, which includes stopping words and stemming. After going through the preprocessing process, the system will look for similarities in user questions with the method model that already contains the question dataset, subject, object, and keyword, where the subject, object, and keyword will pass to the SPARQL query process, where the SPARQL query will give an ontology command to get the appropriate data, after that, it will send the appropriate answer to the user.

3.1 Kingdom of Sumedang Larang

The Sumedang Larang Kingdom was increasingly recognized on the stage of Indonesian national history, especially in Tanah Parahiyangan, after the end of the Pajajaran Kingdom or the Sunda Kingdom in 1580. Sumedang was once revered, admired, and respected, not to mention the poets and poets telling Sumedang stories in their works, whether in titles, fairy tales, chronicles, rhymes, sagas, and even legends.

According to Carita Parahiyangan, the Tembong Agung Kingdom was led by Prabu Guru Haji Aji Putih, as the first king. It is said that Prabu Guru was still related to kinship with Prabu Sri Baduga Maharaja of the Sunda Kingdom at that time. He also had a son named Batara Tuntang Buana who later founded the Sumedang Larang Kingdom. When leading his new kingdom, Batara changed his name to Prabu Resi Tajimalela. Then he said, "Ingsun Medal Ingsun Madangan" (I was born I gave enlightenment), which is now the name SUMEDANG.

3.2 Dataset

For the user's questions and the answers expected by the user to match, we collected two datasets for this system, i.e., the question dataset and the answer dataset

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 3, No. 4, August 2022, Page 322−329
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v3i4.2079

3.2.1 Question Dataset

Based on Table 1 question dataset, we collect historical data from the Sumedang Larang Kingdom. The data consists of a question column and a class column.

Table 1. Question Dataset

No	Question	Class
1	*siapa raja pertama di sumedang larang?* (Who was the first king of sumedang larang?)	*rajapertama*(first king)
2	*Raja Pertama Sumedang larang?* (The First King of Sumedang Larang?)	*rajapertama*(first king)
3	*siapa nama ayah dari batara tuntang buana?* (What is the name of the father of Batara Tuntang Buana?)	*ayahrajapertama* (First king father)
4	*nama ayah batara tuntang buana?* (Batara Tuntang Buana's father's name?)	*ayahrajapertama* (First king father)
5	*siapa nama lain batara tuntang buana?* (What is the other name of Batara Tuntang Buana?)	*namalainrajapertama* (Other name firstking)
6	*nama lain batara tuntang buana?* (Another name for Batara Tuntang Buana?)	*namalainrajapertama* (Other name firstking)
7	*siapa nama istri batara tuntang buana?* (what is the name of Batara Tuntang Buana's wife?)	*istrirajapertama* (first wife)
8	*nama istri batara tuntang buana?* (The name of Batara Tuntang Buana's wife?)	*istrirajapertama* (first wife)
9	*nama anak kembar raja batara tuntang buana?* (the name of the twins of Raja Batara Tuntang Buana?)	*anakrajapertama* (the first child)
10	*siapa nama anak kembar raja pertama sumedang larang?* (what is the name of the twins of the first king of sumedang ban?)	*anakrajapertama* (the first child)

3.2.1 Answer Dataset

Based on Table 2 answer dataset, we collect based on the answers to each question in the question dataset consisting of subject, object, and keywords.

Table 2. Answer Dataset

rajapertama **(the first king)**	***ayahrajapertama*** **(father the first king**	***namalainrajapertama*** **(another name the first king)**	***istrirajapertama*** **(the first king's wife)**	***anakrajapertama*** **(the first king's son)**
pemimpin nama raja pertama (first king name leader)	*pemimpin anak_dari batara tuntang buana* (leader of the children of Batara Tuntang Buana)	*pemimpin nama batara tuntang buana* (leader of the name Batara Tuntang Buana)	*istri batara tuntang buana anak_dari batara tuntang buana* (Batara's wife tuntang Buana's son from Batara Tuntang Buana)	pemimpin *ayah_dari batara tuntang buana* (the leader of the father of Batara Tuntang Buana)
pemimpin nama raja pertama (first king name leader)	*pemimpin anak_dari batara tuntang buana* (leader of the children of Batara Tuntang Buana)	*pemimpin nama batara tuntang buana* (leader of the name Batara Tuntang Buana)	*Istri batara tuntang buana anak_dari batara tuntang buana* (Batara's wife tuntang Buana's son from Batara Tuntang Buana)	*pemimpin ayah_dari batara tuntang buana* (the leader of the father of Batara Tuntang Buana)

3.3 Preprocessing

Preprocessing stage is the stage to convert text into structured form. At this stage, stop word is performed to remove stop word in a text and stemming is to map or restore words to their basic form [12]. Stop words are words that are not processed into the algorithm such as which, to, in etc.

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 3, No. 4, August 2022, Page 322−329
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v3i4.2079

3.3.1 Stopword

In the stopword process, the interrogative sentence will be checked whether the sentence contains words that often appear or not. If there is a word that appears frequently and has no meaning, then the word will be removed from the sentence, as in the three sentences in table 3.

Table 3. Stopword Process

Kalimat Pertanyaan	***Hasil Proses* Stopword**
siapa raja pertama di sumedang larang? (Who was the first king in sumedang Larang?)	*raja pertama di sumedang larang?* (The first king in sumedang Larang?)
sumedang Larang dipimpin pertama *kali oleh siapa?* (Sumedang Larang to be led first by whom?)	*sumedang Larang dipimpin pertama kali oleh?* (sumedang Larang was led by?)
raja siapa yang memindahkan ibukota ke ciguling? (Whose king moved the capital to ciguling?)	*raja yang memindahkan ibukota ke ciguling?* (The king who moved the capital to ciguling?)

3.3.2 Stemming

Stemming is the process of removing the inflection of a word to its basic form, but the basic setup does not mean the same as the root word.

Table 4. Stemming Process

Kalimat Pertanyaan	***Hasil Proses* Steeming**
dimana ibukota sumedang larang dipindahkan pada masa raja gajah agung? (Where was the capital of Sumedang Larang to be moved during the era of the great elephant king?)	*dimana ibukota sumedang larang pindah pada masa raja gajah agung?* (Where was the capital of Sumedang Larang to move during the era of the Great Elephant King?)
dipindahkan kemana ibukota sumedang larang pada masa raja gajah agung? (Where was the capital of Sumedang forbidden during the era of the Great Elephant King?)	*pindah kemana ibukota sumedang larang pada masa raja gajah agung?* (Where was the capital of Sumedang Larang during the era of the Great Elephant King?)
siapakah raja yang memimpin sumedang larang ketika ibukota di ciguling? (who was the king who led sumedang Larang when the capital was in ciguling?)	*siapa raja yang memimpin sumedang larang ketika ibukota di ciguling?* (who was the king who led sumedang larang when the capital was in ciguling?)

3.3.3 Lowercase

Lowercase is converting all capital letters in a sentence to lowercase. The following is an example of the lowercase process in a sentence.

Table 5. Lowercase Process

Kalimat Pertanyaan	***Hasil Proses* Lowercase**
Raja Pertama Sumedang Larang Adalah? (The First King of Sumedang Larang Is?)	*raja pertama sumedang larang adalah?* (the first king of sumedang Larang is?)
Siapa nama Anak Raja Pertama Sumedang larang? (What is the name of the Son of the First King of Sumedang Larang?)	*siapa nama anak raja pertama sumedang larang?* (what is the name of the son of the first king of sumedang Larang?)
Pada Tahun Berapa Geusan Ulun memimpin Kerajaan Sumedang Larang? (In what year did Geusan Ulun lead the Sumedang Larang Kingdom?)	*pada tahun berapa geusan ulun memimpin kerajaan sumedang larang?* (In what year did Geusan Ulun lead the Sumedang Larang Kingdom?)

3.4 TF-IDF Process

TfidfVectorizer is the word Inverse Domain Frequency (IDF), and Term Frequency (TF) is the calculation of the corresponding TF-IDF value. We use the TF-IDF score to weigh words in sentiment analysis. And IDF for the word w in the text and TF-IDF for the word w in the question. We perform the calculations using the following formula:

$$idf(w) = \log(\frac{total\ number\ of\ questions}{No.of\ questions\ with\ column\ class}) \quad (2)$$

$$Tf_{idf(w,d)} = tf(w,d) * idf(w) \quad (3)$$

Where $tf(w,d)$ is the word w in the document d or the class of the column, or the frequency term in the question column, such as the number of times w appears in document d divided by the total number of words in

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 3, No. 4, August 2022, Page 322−329
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v3i4.2079

the document. Finally, the corpus data vector will divide into two parts, i.e., the training set and the test set with a ratio of 8:2 and 7:3 [14].

3.5 Query SPARQL

SPARQL stands for Simple Protocol and RDF Query Language. The SPARQL standard defines a network protocol for exchanging requests and a language for expressing requests. SPARQL uses a SQL-like syntax to represent queries. Our research uses the SPARQL query. For example, we can see Figure 2.

```
SELECT *
WHERE { ?subject    rdf:type    sumedanglarang:"""+subject+""" ;
    sumedanglarang:kata_kunci    ?katakunci ;
    sumedanglarang:"""+object+"""  ?object
      FILTER(STR(?katakunci) = """+katakunci+""")
      }
```

Figure 2. SPARQL Query our System

The first line is to read the class of the ontology. The second line is to get the individual classes in each class. The third line is to get the individual classes that belong to the keyword of each class. 4th line to filter the data that matches the specified keyword.

3.6 Ontology

Ontology is a philosophy initially described as the science of being. However, in the language of modern computer science, ontology is described as the essence of concepts and their relationships—knowledge of ontologies that are very useful for question classification systems, question analyzers, and answer extractors.

We perform the ontology knowledge base design semi-automatically. First, corpus-based information will be extracted into several terms, concepts, attributes, relationships, and examples. The ontology elements are then imported into the project and manipulated manually [15]. In our research, we use relation on ontology. Figure 3 shows a part of our research.

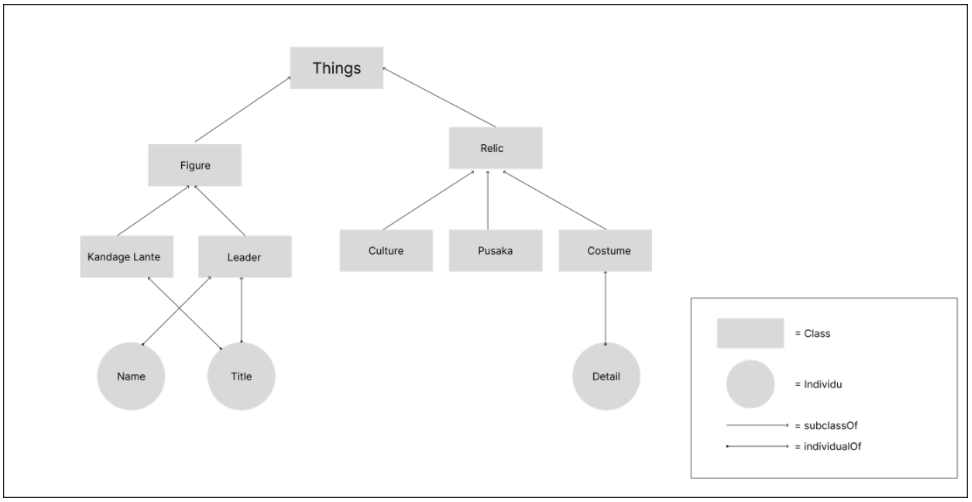

Figure 3. Schema Ontology

3.7 Question Answering (QA)

QA system is a system for enabling computers understand the questions asked by using natural language other than that QA system can respond to questions based on available information [9]. QA systems have special significance and advantages over machines search [10]. In QA, users simply type a question and display the list of web pages from the most suitable [11]. Users and analysts want targeted answers to their questions in the absence of addition of unrelated information [11]. So, the answer you get closely related to the question and in accordance with the expected answer desired.

3.8 Naïve Bayes

Naïve Bayes Classifier is a classification method that uses Bayes' theorem [16]. If there are two separate events (A and B), then Bayes' theorem can be formulated as follows [16]:

$$P(A|B) = \frac{P(A)}{P(B)} P(B|A) \quad (4)$$

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 3, No. 4, August 2022, Page 322−329
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v3i4.2079

The Naive Bayes method assumes that each question data does not depend on other question data. The Naive Bayes method can provide satisfactory results when applied in the field of text classification [17]. In this study, we use Naive Bayes Multinomial because it is often used in text classification [18]. Multinomial Naive Bayes is a supervised learning method in which each question data requires a label before training. The probability that a document d is in label a can be calculated using the following formula [18].

$$P(a|d) \propto P(a)\Pi_{k=1}^{n} P(t_k|a) \quad (5)$$

$P(a|d)$: The probability of document d being in label a
$P(a)$: Prior probability of a document being in label a
$\{t_1, t_1, .., t_n\}$: The token in document d which is part of the vocabulary with the number n
$P(t_k|a)$: The conditional probability of the term t_k being in the document on the label a

3.9 Evaluation

The accuracy test uses a question dataset which has a category column and an answer column. The category column is labeled encoder and then divided into 8:2 and 7:3 ratio. Then the result of the division is in the form of text which will be converted into a number, where the number will form a vector using the TF-IDF method. (Dalam Penelitian ini, kami melibatkan decision tree sebagai base model). In the classification process, we use Naive Bayes Multinomial. So, the result of the test is shown in figure 4.

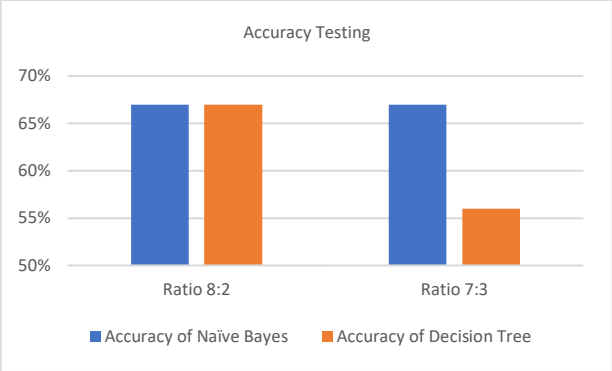

Figure 4. Accuracy Testing

Based on figure 4, the most appropriate accuracy for this study is the one using the Naive Bayes method where the two ratios, such as 8:2 and 7:3 have a stable accuracy rate of 67%. This is because the Naive Bayes method assumes that each question data does not depend on other question data so that Naive Bayes is able to give pretty good results when applied in the field of text classification where this research uses the Naive Bayes method or Decision Tree to classify questions and answers in the form of text.

We notice that there are several factors that affect the results of accuracy testing, the similarity of the question sentence because when there are one or two words that are the same between one question sentence with another, there will be ambiguity in the classification process. In addition, the number of questions also affects accuracy because the more questions, the more similar the question sentence. Meanwhile, in ratio of 8:2, the Decision Tree has accuracy of 67% and in ratio 7:3, it has accuracy of 56%. Both results are unstable and smaller than the results of Naive Bayes.

4. CONCLUSION

This study uses the Naive Bayes method on the Question Answering system using an ontology, where the question dataset will be categorized according to the answer data that has been collected according to our information and insights about the Sumedang Larang Kingdom. And the results of testing the accuracy of each method include the Naive Bayes method. With higher accuracy, 8:2 67% ratio accuracy, and 7:3 67% ratio, the Decision Tree accuracy is only 56%. This is due to the classification process of each method. In the Naïve Bayes method, the classification process assumes that each question data does not depend on each other, while the Decision Tree method assumes that each question data depends on each other, so it can be concluded that in this study the most appropriate method to use is the Naive Bayes method for answer questions from users about the Sumedang Larang Kingdom.

REFERENCES

[1] M. Sarrouti and S. Ouatik El Alaoui, “A Biomedical Question Answering System in BioASQ 2017,” in *BioNLP 2017*, 2017, pp. 296–301. doi: 10.18653/v1/W17-2337.

ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 3, No. 4, August 2022, Page 322−329
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v3i4.2079

[2] Z. K. A. Baizal, D. H. Widyantoro, and N. U. Maulidevi, "Computational model for generating interactions in conversational recommender system based on product functional requirements," *Data & Knowledge Engineering*, vol. 128, p. 101813, Jul. 2020, doi: 10.1016/j.datak.2020.101813.

[3] S. F. Kusuma, D. O. Siahaan, and C. Fatichah, "Automatic Question Generation In Education Domain Based On Ontology," in *2020 International Conference on Computer Engineering, Network, and Intelligent Multimedia (CENIM)*, Nov. 2020, pp. 251–256. doi: 10.1109/CENIM51130.2020.9297991.

[4] M. S. Ayundhita, Z. K. A. Baizal, and Y. Sibaroni, "Ontology-based conversational recommender system for recommending laptop," *Journal of Physics: Conference Series*, vol. 1192, p. 012020, Mar. 2019, doi: 10.1088/1742-6596/1192/1/012020.

[5] D. Baeva and D. Atanasova, "Ontology based resource for history education," *TEM Journal*, vol. 7, no. 4, pp. 782–786, 2018, doi: 10.18421/TEM74-13.

[6] R. Cahya Hutama and R. Titi Komalasari, "STRING (Satuan Tulisan Riset dan Inovasi Teknologi) APLIKASI CHATBOT BERBASIS TEKS MENGGUNAKAN ALGORITMA NAIVE BAYES CLASSIFIER FAQ GRABADS."

[7] A. Abdi, N. Idris, and Z. Ahmad, "QAPD: an ontology-based question answering system in the physics domain," *Soft Computing*, vol. 22, no. 1, pp. 213–230, Jan. 2018, doi: 10.1007/s00500-016-2328-2.

[8] M. Rani, A. K. Dhar, and O. P. Vyas, "Semi-automatic terminology ontology learning based on topic modeling," *Engineering Applications of Artificial Intelligence*, vol. 63, pp. 108–125, Aug. 2017, doi: 10.1016/j.engappai.2017.05.006.

[9] A. Trotman, S. Geva, and J. Kamps, "Report on the SIGIR 2007 workshop on focused retrieval," *ACM SIGIR Forum*, vol. 41, no. 2, pp. 97–103, Dec. 2007, doi: 10.1145/1328964.1328981.

[10] A. Bouziane, D. Bouchiha, N. Doumi, and M. Malki, "Question Answering Systems: Survey and Trends," *Procedia Computer Science*, vol. 73, pp. 366–375, 2015, doi: 10.1016/j.procs.2015.12.005.

[11] JO and Taeho, *ext mining: Concepts, implementation, and big data challenge*. 2018.

[12] V. Kumar and B. Subba, "A TfidfVectorizer and SVM based sentiment analysis framework for text data corpus," in *2020 National Conference on Communications (NCC)*, Feb. 2020, pp. 1–6. doi: 10.1109/NCC48643.2020.9056085.

[13] S. Kiv, Y. Wautelet, S. Heng, and M. Kolp, "OBAMA, an Ontology-Based Software Tool for Agile Method Adoption," Jun. 2022, [Online]. Available: http://arxiv.org/abs/2206.02207

[14] J. Fu, J. Xu, and K. Jia, "Domain Ontology Based Automatic Question Answering," in *2009 International Conference on Computer Engineering and Technology*, Jan. 2009, pp. 346–349. doi: 10.1109/ICCET.2009.132.

[15] "Jurnal Matrik".

[16] M. A. Rosid, A. S. Fitrani, I. R. I. Astutik, N. I. Mulloh, and H. A. Gozali, "Improving Text Preprocessing for Student Complaint Document Classification Using Sastrawi," in *IOP Conference Series: Materials Science and Engineering*, Jul. 2020, vol. 874, no. 1. doi: 10.1088/1757-899X/874/1/012017.

[17] C. D. Manning, Prabhakar. Raghavan, and Hinrich. Schütze, *Introduction to information retrieval*. Cambridge University Press, 2008.

[18] K. Khalifa and N. Omar, "A HYBRID METHOD USING LEXICON-BASED APPROACH AND NAIVE BAYES CLASSIFIER FOR ARABIC OPINION QUESTION ANSWERING," *Journal of Computer Science*, vol. 10, no. 10, pp. 1961–1968, Oct. 2014, doi: 10.3844/jcssp.2014.1961.1968.