International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys.13, No.1 (Apr-23)

http://dx.doi.org/10.12785/ijcds/130188

Public Hospital Review on Map Service with Part of Speech Tagging and Biterm Topic Modeling

Moh Makruf[1] and Arif Bramantoro[2]

[1] *Computer Science Postgraduate Program, Universitas Budi Luhur, Jakarta, Indonesia*
[2] *School of Computing and Informatics, Universiti Teknologi Brunei, Bandar Seri Begawan, Brunei Darussalam*

Received 27 Apr. 2022, Revised 6 Apr. 2023, Accepted 26 Apr. 2023, Published 30 Apr. 2023

Abstract: Health is a basic human need. Increasing standard of life influences the quality of health. It demands health service providers improve quality better service and provide satisfaction for consumers as health care users. Hospital as one of the means of health services is required to improve the quality of services. The study found that customers tend to choose private hospitals over public hospitals because of service factors. Public hospitals are gradually being abandoned. This is challenging for public hospitals to improve services. One of the scientific techniques to know the level of public satisfaction with health services, especially hospitals, is to examine the reviews from users. In this research, we use public mapping review as a data source. From the review, it can be known what topics are mostly discussed specifically for various ratings. A suitable model is required to find out the topics in the review that have a low rating so that it can be used as a suggestion for improving health services. Topic modeling is best achieved through the text mining method. The study proposed the use biterm topic model suitable for short text in the reviews of the map platform. Short text reviews are characterized by sparse data, the small number of words that appear as topic builders, and rare word contexts related to the topic. The result shows that biterm topic model can produce topics with an acceptable combination of accuracy and performance. The model with the addition of part-of-speech tagging on noun-only is unable to increase the accuracy of the model, compared to the one without the addition of part-of-speech tag. However, the addition of part-of-speech tagging on noun-only can improve the performance of the model.

Keywords: Health Services, Topic Model, Part-of-Speech, Biterm

1. Introduction

Healthcare service is essential for human life. The increasing standard of living influences the demand for better health quality. This requires healthcare service providers to alleviate the service quality standards and provide negligible satisfaction for consumers as health service users [1]. Hospitals as healthcare facilities and the institutions providing health services, need to maintain the user centered quality of service [2] that is always evaluated.

One study found that most people tend to choose private hospital services [3], which confirms the common sense amongst most patients. Thus, it is estimated that government-owned hospitals will gradually be abandoned. It is a challenge for public hospitals to improve services to avoid this prediction.

One of the common ways to find out the level of public satisfaction with health services in hospitals is to use a public mapping review. This is because the public map service gains more and more attention from users to provide their reviews. The review provides the topics that are mostly talked about and the selection of reviews with low ratings, i.e. from one to three. This research proposes to use the biterm topic model (BTM) [4] algorithm to extract topics on short reviews in map services regarding hospitals. The mapping review is a kind of systematic review. The systematic review has strict requirements for stages, steps and procedures [5]. The BTM algorithm is suitable for modeling short text topics according to the authors in [6]. The main characteristic of short text documents is data sparsity that leads to incomplete information. It also leads to the fact that the number of words that build the topic is usually only one, and there are few words that are related to the topic.

We propose to use part-of-speech tagging for noun-only which aims to improve the performance and accuracy of the BTM model. We compare the use of noun-only and all part-of-speech tags to provide the best model in term of accuracy and performance. Python is chosen as a modeling and implementation language to improve the flexibility and visualization of analysis due to the wide availability of its libraries.

E-mail address: 1911601175@student.budiluhur.ac.id, arif.bramantoro@utb.edu.bn

The paper is structured as follows. Section 1 introduces the problem of topic modeling. Section 2 presents the literature review. Section 3 presents the proposed methodology. Section 4 presents the data preprocessing and implementation. Section 5 presents the result and discussion. Finally, section 6 provides the conclusion.

2. Literature Review

Topic modeling is gaining more interest amongst researchers and practitioners following the popularity of machine learning and natural language processing techniques. It is argued that the more techniques combined as composite components the better the result obtained [7]. The previous research conducted by [8] on topic modeling using latent Dirichlet allocation (LDA) [9] and part-of-speech tagging concluded that the topic modeling with the reduction of the corpus by including only nouns can produce an equivalent or even increased topic semantic coherence compared to the one with the complete corpus. A similar empirical result for topic modeling using LDA and part-of-speech tagging was provided by another study in [10] that shows the increase of performance by accommodating semantic and syntactic information between two text documents.

Research by [11] concluded that the BTM method is the best and most stable based on all coherence measures, whereas the non-specific short text LDA model showed not suitable without additional data pre-processing mainly due to the data availability. Research by [12] concluded that BTM is somewhat better than LDA and word network topic model but worse than the topic keyword model for any classification cases except for the book reviews data set, which is characterized by relatively short documents that seem to be better handled by BTM. Research by [13] using BTM and K-Means concluded that experimental results on a collection of short texts available on microblogs indicated that this method could model topics effectively.

It is interesting to note that there are several variants of BTM technique have been researched. A study by [14] showed that a robust user sentiment biterm topic mixture model extracted topics more coherent and informative. A study by [15] experimented with a comparison between clustering methods showing that short text modeling and an improved single-pass algorithm using BTM increased clustering efficiency. It was argued that this combination could effectively solve the problem of data sparsity in short texts. [16] proposed relational BTM to model short text using word embeddings. The short text is scraped from 5.42 million documents on 2011 Text Retrieval Conference microblog track and 3.37 million documents on Kaggle. The analysis focuses on the question title variable that results in the superiority of the algorithm compared to basic topic modeling techniques. Authors in [17] found that the proposed FastBTM that exploits hashtags and aliases on 16 million of 2011 tweets, 142,627 Yahoo answers, and Enron emails from 150 users was able to reduce the complexity of sampling during the BTM topic modeling from $O(k)$ to $O(1)$.

Those previous works give us the motivation to implement the BTM algorithm to extract topics from short documents using the dataset from the mapping review that are not yet presented in previous research. Moreover, this research proposes to add part-of-speech tag to improve the performance and accuracy of the model.

3. Methodology

BTM is a topic modeling technique that is generally used for short text. The main idea behind the use of BTM is to study the topic from documents that mostly contain short text based on aggregated biterms throughout the corpus to address sparsity issues in a single document. In particular, the entire corpus is considered a mixture of topics, in which each biterm is taken from a particular topic independently [4]. Although there is another method of topic modeling on short texts, such as LDA, it has been paved in the literature review that BTM is superior. Hence, it is trivial to compare BTM to other methods.

In most topic modeling methods, topics are represented as interconnected groups of words where the relationship is based on the pattern of word emergence. Conventional topic modeling technique exploits word emergence patterns to express the latent semantic structure of the corpus implicitly by modeling word generation from each document. This approach is very sensitive to the length of the document because the pattern of word emergence in short documents is incredibly rare and unreliable. If all the patterns of word emergence on the corpus are combined, the frequency becomes more stable and the correlation between words becomes clearer. This is based on the underlying idea of developing the BTM method.

According to [4], the first stage in the BTM method is biterm extraction. A biterm is a pair of non-sequential words that appear on a document. For example, a short document with four words, $w1, w2, w3$, and $w4$ combinatorially form three biterms namely $(w1, w2), (w1, w3), (w1, w4)(w2, w3), (w2, w4)$, and $(w3, w4)$.

Random initialization of the topic is given to each biterm as the first step in this method. Furthermore, parameter estimation is also required for the BTM method. This is the process of sampling to obtain the value of the probability distribution of the word on a particular topic and the probability distribution of the topic on the document.

Output parameters estimation is obtained by a sampling probability distribution. More specifically, BTM uses Gibbs sampling [18] which requires an iteration. For each iteration and each topic, random sampling is satisfied based on the conditional probability distribution.

Once the iteration process for Gibbs sampling is complete, BTM calculates the probability distribution of words

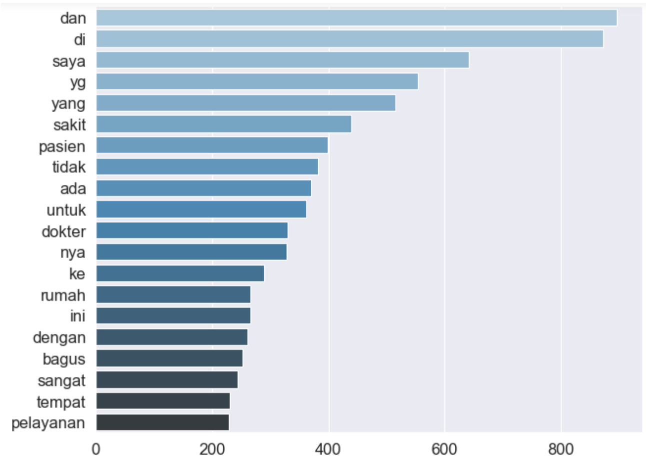

Figure 1. Common Word Frequency

on a topic and the probability distribution of the topic on the document. Each word on the corpus calculates its probability distribution to the topic. Each document calculated the probability distribution of its topic.

4. Implementation

The proposed steps in this research are data preparation, model building, and evaluation. Data preparation is non-trivial but time-consuming task. In this step, data are collected, combined, transformed, and cleaned. In detail, it includes checking for the completeness of the content, looking at the dimension, reviewing the structure of the input dataset, peeking into data, and checking for any missing data based on a complete review. Python is chosen to be a modeling language because it provides robust libraries required in this research.

Dataset is gathered from the mapping review of Indonesian public hospitals which consists of 23 variables and 2,115 rows. Due to its high popularity in the region, the review from public map service platform is chosen as a platform in this study. Variable *review_text* is used for modeling and *review_rating* is used for selection. During data preparation, we use exploratory data analysis to help observe data before continuing to the next step.

Figure 1 shows the words with the greatest frequency, including the common words, such as *dan* (and), *di* (in), *saya* (me), *yang* (who), *tidak* (not), and others. Some of these words are part of the stopwords that will be eliminated in the next process. The top two words with a frequency of over 800 are the words *dan* (and) and *di* (in). To understand better the frequency of the common words, the word cloud visualization is presented in Figure 2.

We can examine the number of characters in each review by using a histogram. This can give us a rough idea of the length of the review. Figure 3 shows that the review consists of from zero to 2,000 characters, with an average of fewer than 300 characters. This is because short text tends to have

Figure 2. Common Word Cloud

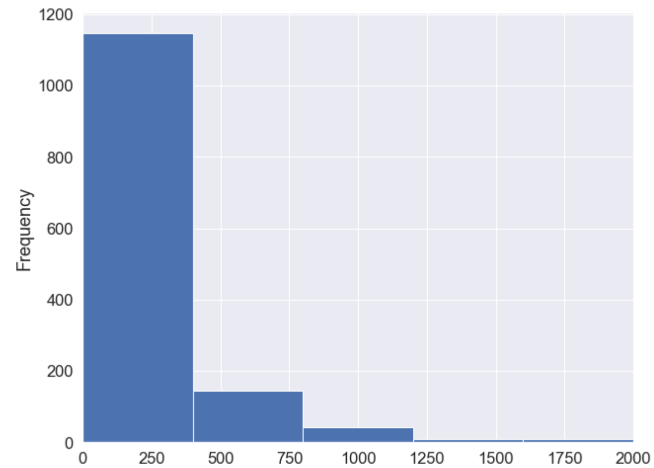

Figure 3. Number of Characters in Each Review

fewer character in total.

To explore the review in more detail, it is necessary to examine it at the word level. The number of words that appear in each review ranges from zero to 200 words, with the dominant number between zero and five words as shown in Figure 4. This is because short text tends to have few words in each sentence.

It is also important to examine the length of each word in the review. The word length of the review is ranging from two to 12 characters, with the most common length being five characters as shown in Figure 5. This is because short text tends to have short words on average. However, it is interesting to note that not so many words with a length of fewer than two characters. This can be inferred that there are not many slang words used in the review.

Before the data cleaning process, the selection of reviews that have a low rating (from one to three out of five) is selected. In total, we obtained 855 reviews. The data cleaning stage begins with the deletion of null, NA, or blank data. After the deletion of these data, the total number of datasets became 629 reviews.

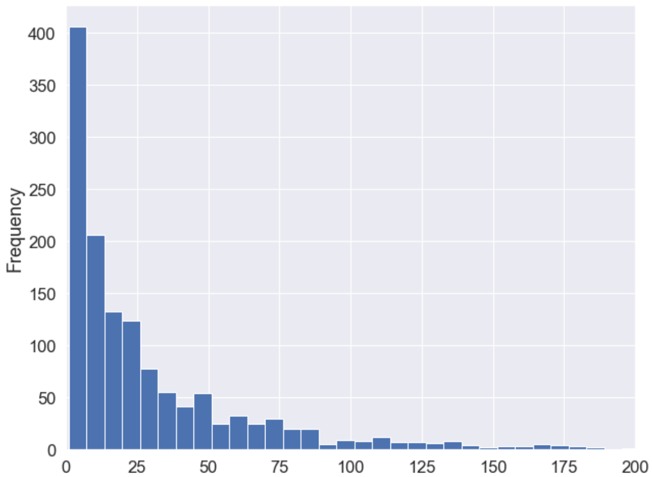

Figure 4. Number of Words in Each Review

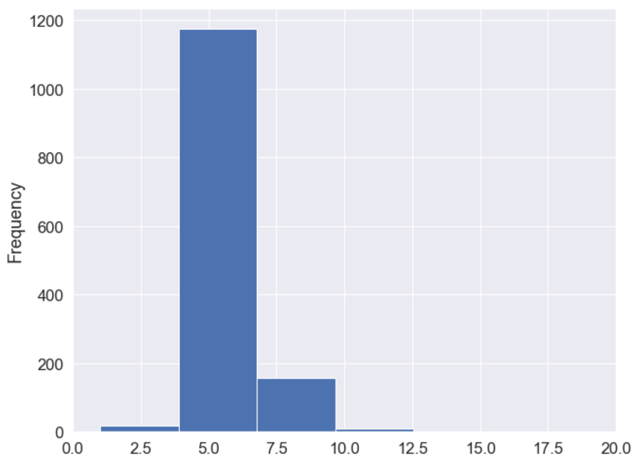

Figure 5. The Average Length of Character in a Word

The next step is to remove emojis and eliminate punctuation. It includes dots, commas, HTML characters, punctuations in the form of other characters, and numbers. The next process is converting all characters into lowercase characters as required by most machine learning algorithms.

In the dataset, there are original sentences in English. Since the focus of this research is the review in the Indonesian language, the original English version of the sentences is omitted. It also aims to prevent duplication of data derived from the original sentence that is automatically translated by a translation service as part of the feature in the review platform.

The model preprocessing process consists of stemming/lemmatization, tokenization, stopword removal, and part-of-speech tag. We use customized stemming by eliminating the suffix of the word. The tokenization is implemented by using the $word_tokenize$ function in Python's NLTK module, and stopword removal is implemented by

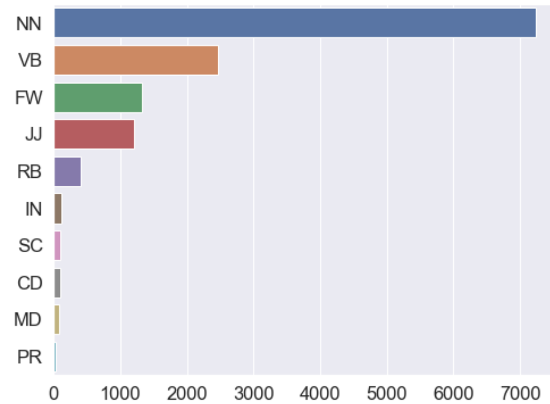

Figure 6. Part of Speech Tagging Frequency

using nltk.corpus.stopwords function in the Indonesian language.

To improve the performance, there are 181 stopwords added in this study based on personal justification as a native of the Indonesian language. In the part-of-speech tagging process, the library is NLTK CRFTagger set up for Indonesian tagger namely pos-tagger-indonesia-model.tagger. There is a need for manual adjustment on the results of part-of-speech tagging, because several words were identified with the wrong part-of-speech tag. This intervention is unavoidable because of the lack of language resources for Indonesian.

There is also a need to have a comprehensive illustration of the result of part-of-speech tagging. In Figure 6, it can be seen that nouns (NN) dominated the review followed by verbs (VB) and adjectives (JJ). It can be inferred that the focus of this research should be on NN, although the use of the VB is quite promising, especially with the addition of the dataset in the future. Moreover, it is important to investigate further which NN mostly appears in the reviews. Figure 7 shows that the words pasien (patient), dokter (doctor), and pelayanan (service) are dominant in the review.

The Biterm algorithm is implemented in the C++ programming language and several parts of the algorithm are in Python programming language. A different programming language is required because there is a need to completely execute in a shell script to support the trial-and-error experiments.

For evaluating and visualizing the models, the authors utilized several Python libraries, such as pandas, nltk, numpy, matplotlib, scipy, and tmplot. Each library has various roles in producing the best analysis results. It is also important to note that the libraries need the latest version to be installed.

The first implementation is the setting of the hyperpa-

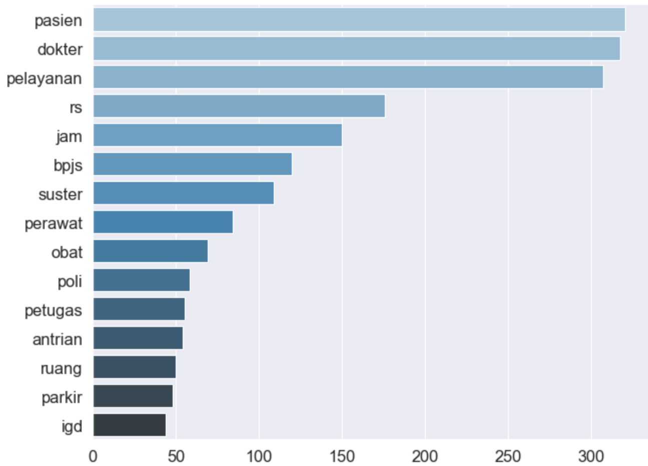

Figure 7. Word Frequency on Noun-Only Tagging Frequency

rameter by selecting a set of optimal parameters for algorithmic learning. The hyperparameters used in this study are alpha-beta values, number of iterations, and number of topics. The model is created on top of the selected datasets with noun-only tags and the overall datasets with all part-of-speech tags.

Before training the model, it is necessary to set two parameters α and β. Due to the large complexity of BTM processing time, we use the parameter settings recommended by [4], i.e. $\alpha = 50/k$ and $\beta = 0.01$. We proposed to add two β values of 0.001 and 0.05 for experimental trial and error with a fixed α value of $50/k$.

The next step is the calculation of the number of topics. It is necessary to ascertain the convergence of models for all parameter settings. By calculating log-likelihood, we can get the amount required for Gibbs sampling iterations. Hence, it requires checking the convergence of each log-likelihood model.

The number of topics K is initialized with the values of 5, 10, 15, and 20, while the number of iterations is initialized with the values from 10 to 200 with the multiplication of 10. The initialization process requires several experiments to obtain these values.

By using parameters $\alpha = 50/k$ and $\beta = 0.01$ for noun-only dataset, the perplexity value has reached a stable condition at the 100^{th} iteration as illustrated in Figure 8, Therefore, 100 can be accepted as the suitable number of iterations. With the same reasoning, it is found that the perplexity value has reached a stable condition for all part-of-speech tag dataset at the 100^{th} iteration as illustrated in Figure 9.

By using parameters $\alpha = 50/k$ and $\beta = 0.001$ for noun-only dataset, the perplexity value has reached a stable condition at the 110^{th} iteration as illustrated in Figure 10. Therefore, 110 is considered a suitable number of iterations.

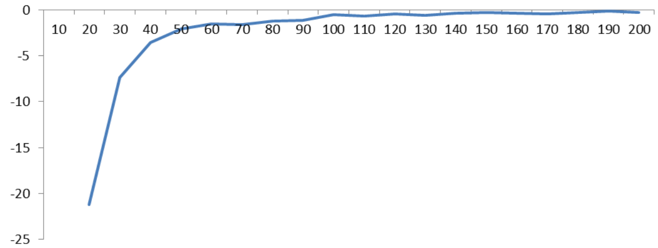

Figure 8. Perplexity Trend with $\alpha = 50/k$ and $\beta = 0.01$ for Noun-Only

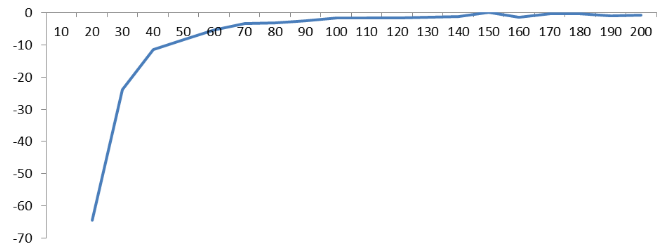

Figure 9. Perplexity Trend with $\alpha = 50/k$ and $\beta = 0.01$ for All Part-of-Speech

For all part-of-speech tag dataset, the perplexity value has also reached a stable condition at the 110^{th} iteration as illustrated in Figure 11.

By using parameters $\alpha = 50/k$ and $\beta = 0.05$, for noun-only dataset, the perplexity value has reached a stable condition at the 110^{th} iteration as illustrated in Figure 12. Therefore, 110 is considered a suitable number of iterations. For all part-of-speech tag dataset, however, the perplexity value has reached a stable condition at the 100^{th} iteration as illustrated in Figure 13. Hence, it is interesting to highlight the difference in this case.

Once the number of iterations is determined, the next process is to calculate the number of topics. Conducting several experiments to determine the number of topics is an important stage to produce the best model. A good model is one with a low perplexity value. In other words, the lower the perplexity value the better the model accuracy.

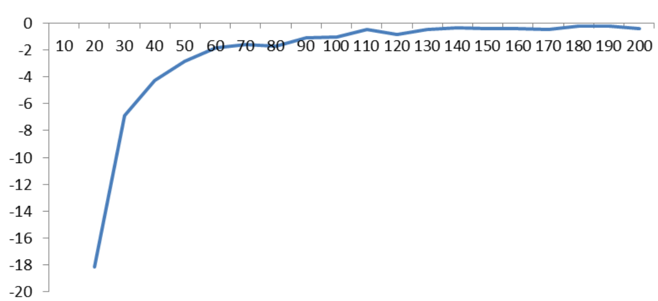

Figure 10. Perplexity Trend with $\alpha = 50/k$ and $\beta = 0.001$ for Noun-Only

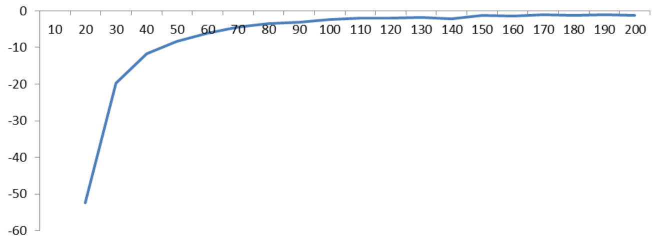

Figure 11. Perplexity Trend with $\alpha = 50/k$ and $\beta = 0.001$ for All Part-of-Speech

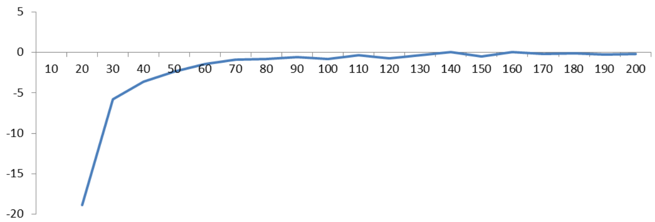

Figure 12. Perplexity Trend with $\alpha = 50/k$ and $\beta = 0.05$ for Noun-Only

We determine the number of topics by arbitrarily applying an initial minimum value of two and a maximum value of ten topics. These numbers of topics are selected because the datasets are relatively small, and the documents are relatively short.

If more topics are selected, there is a possibility of a repeated occurrence of several words within the same topic. Hence, the evaluation metric is required to measure how good the topics can be interpreted by human. This paper proposes to use coherence score as the evaluation metric. Based on the experimental trials the emerging coherence value is recorded for analyzing its value trend. Therefore, the number of selected topics is the number of topics with the highest coherence score at the end of the process.

By obtaining the values of α, β, and the number of iterations from the previous step, the maximum coherence score for noun-only is -44.27 with two topics. On the other hand, for all part-of-speech tag dataset with the specified values of α, β, and the number of iterations obtained from the previous step, the maximum coherence score is -43.47 with the number of topics being two. A summary of all

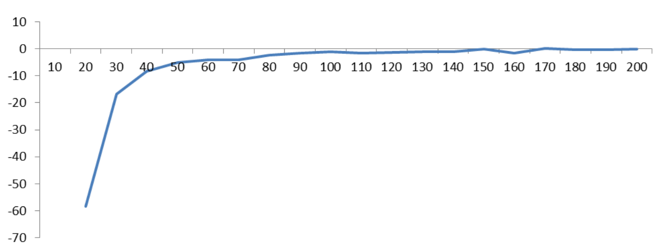

Figure 13. Perplexity trend with $\alpha = 50/k$ and $\beta = 0.05$ for all part-of-speech

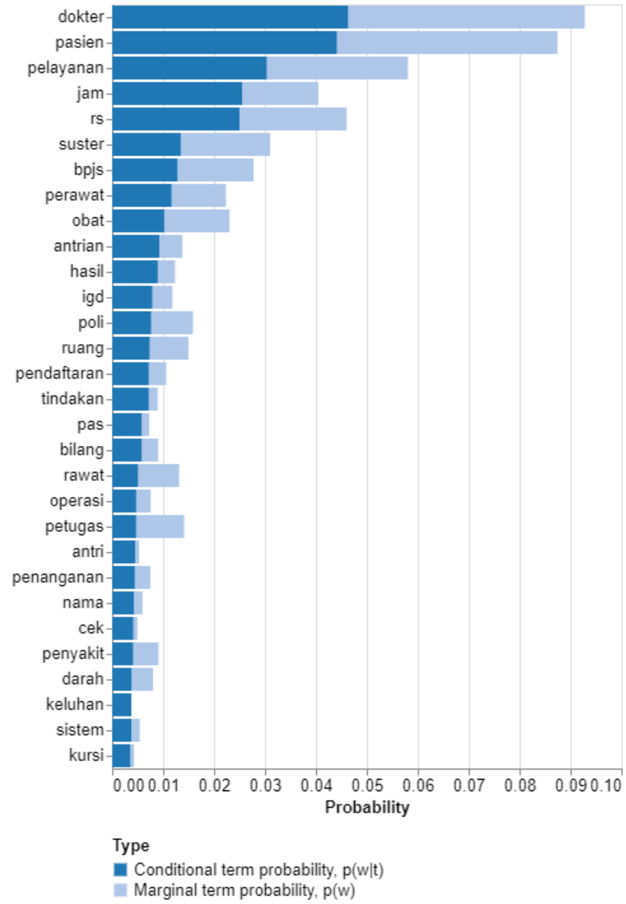

Figure 14. Probability of Words for Noun Only in Topic 0

parameters used in the model can be seen in Table I.

5. Results and Discussion

Once the parameters are determined to get the most out of the topic, the next process is to look at the results of topic modeling. In the noun-only dataset, there are ten most relevant words identified by the probability of words in each topic as illustrated in Figure 14. Topic 0 consists of the words *dokter* (doctor), *pasien* (patient), *pelayanan* (service), *jam* (hour), *rs* (hospital), *suster* (practical nurse), *bpjs* (national insurance), *perawat* (nurse), *obat* (medicine), *antrian* (queue).

For Topic 1, the result is shown in Figure 15. It consists of the words *dokter* (doctor), *pasien* (patient), *pelayanan* (service), *rs* (hospital), *suster* (practical nurse), *jam* (hour), *bpjs* (national insurance), *obat* (medicine), *parkir* (park), *perawat* (nurse).

To grasp the overall trend, the topic distribution is required. The topics distribution in the noun-only dataset is 52.89% of documents in topic 0 and 47.11% of documents in topic 1 as shown in Figure 16.

TABLE I. Model Parameters

Part-of-Speech	α	β	Iteration Number	Coherence Score	Topic Number
Noun only	50/k	0.01	100	-44.27	2
		0.001	110	-44.31	2
		0.05	110	-44.45	2
All part-of-speech	50/k	0.01	100	-46.42	2
		0.001	110	-43.47	2
		0.05	100	-46.98	3

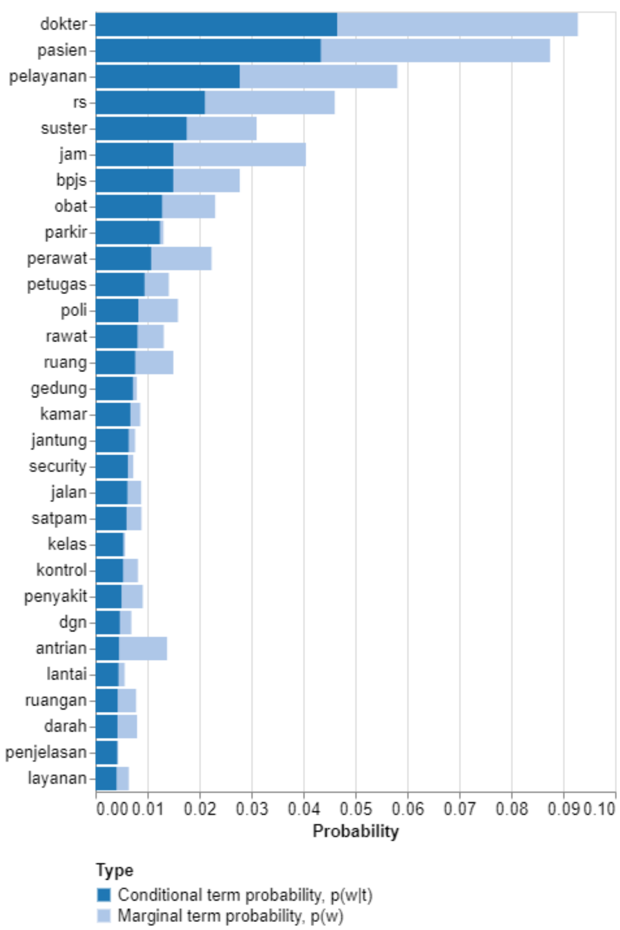

Figure 15. Probability of Words for Noun Only in Topic 1

In the all part-of-speech tag dataset, there are ten most relevant words identified by the probability of words in each topic as illustrated in Figure 17. Topic 0 consists of the words *dokter* (doctor), *pasien* (patient), *jam* (hour), *pelayanan* (service), *rs* (hospital), *bpjs* (national insurance), *suster* (practical nurse), *poli* (clinic), *antrian* (queue), and *obat* (medicine).

For Topic 1, the result is shown in Figure 17. It consists of the words *pasien* (patient), *pelayanan* (service), *dokter* (doctor), *rs* (hospital), *suster* (practical nurse), *perawat* (nurse), *bpjs* (national insurance), *tolong* (assist), *buruk* (low quality), *jam* (hour).

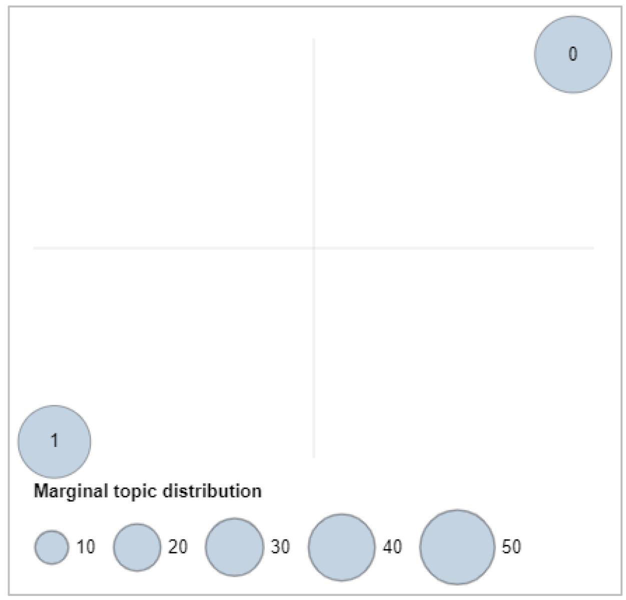

Figure 16. Topic Distribution for Noun Only Dataset

On the other hand, the topics distribution in the all part-of-speech tag dataset is 42.16% of documents in topic 0 and 57.84% of documents in topic 1 as shown in Figure 18.

The model is evaluated through an analysis of previous input parameters to compare the accuracy and performance of models on the noun-only dataset and the dataset of all part-of-speech tags. The first scenario for this comparison is by using the same parameters as in the noun-only and all part-of-speech tag datasets, and the second one is by using the optimal parameters from each noun-only and all part-of-speech tag datasets.

The first scenario uses the same parameter for both datasets i.e. $\alpha = 50/k$, $\beta = [0.01, 0.001, 0.05]$, number of topics = [2,3,. . . ,10], and number of iterations = [10,20,.., 200], it was obtained that the average coherence score for noun-only dataset is -50.31, while average coherence for all part-of-speech tag dataset is -49.03. It means that all part-of-speech tag dataset has a higher coherence score.

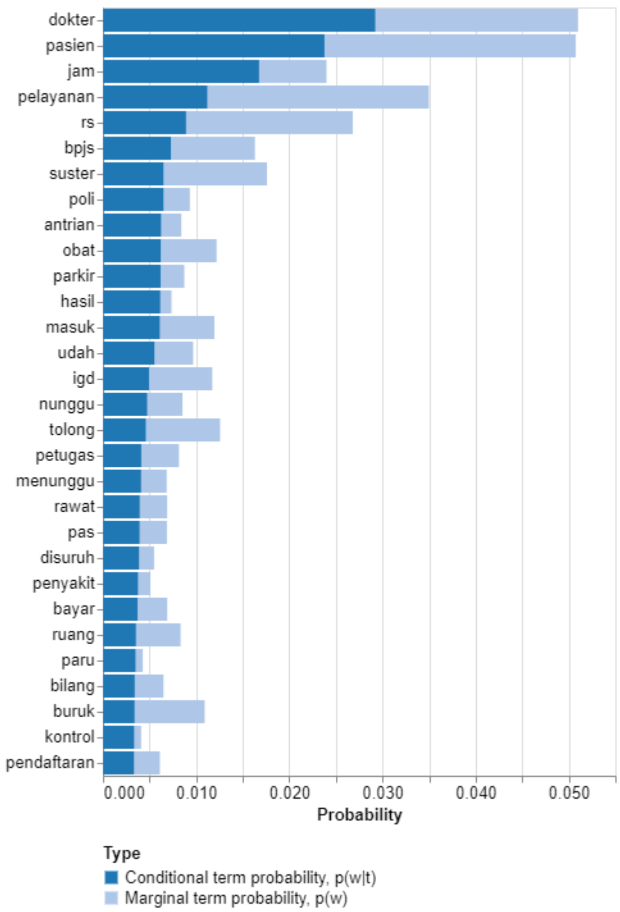

Figure 17. Probability of Words for All Part-of-Speech in Topic 0

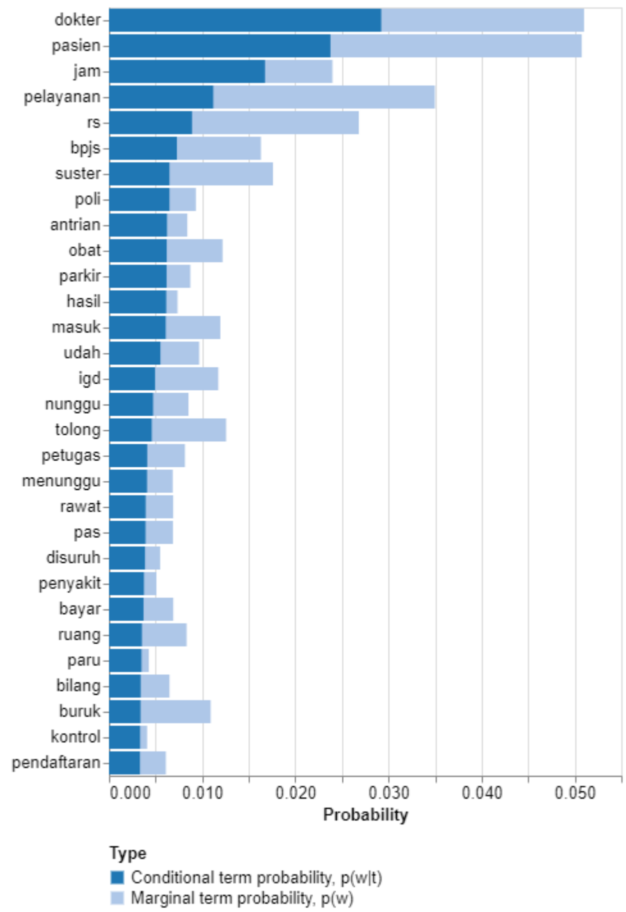

Figure 18. Probability of Words for All Part-of-Speech in Topic 1

TABLE II. The First Scenario of Model Comparison Using the Same Parameters

Metrics	Noun Only	All Part-of-Speech
Model Accuracy (Coherence Score)	-50.31	-49.03
Processing Time (Seconds)	1.49	3.23

TABLE III. The Second Scenario of Model Comparison using Optimal Parameters

Metrics	Noun Only	All Part Of Speech
Model Accuracy (Coherence Score)	-44.27	-43.47
Processing Time (Seconds)	1.01	2.47

The average time required for model creation for noun-only dataset is 1.49 seconds, while the one for all part-of-speech tag dataset is 3.23 seconds. It means that noun-only dataset has a better performance. The first scenario comparison is summarized in Table II.

In the second scenario, the optimal parameters for noun-only dataset are $50/k$ for alpha value, 0.01 for beta value, two for the number of topics, and 100 for the number of iterations. These parameters are considered optimal because the experiment obtained an average coherence score of -44.2713035. For the all part-of-speech tag dataset, the optimal parameters are $50/k$ for the α value, 0.001 for the β value, two for number of topics, 110 for and number of iterations. The use of these parameters was able to achieve the average coherence score of -43.47. It means that all part-of-speech tag dataset has a higher coherence score. For performance analysis, the model creation for noun-only dataset took 1.01 seconds, while the one for all part-of-speech tag dataset took 2.47 seconds. It means that noun-only dataset has a better performance. A summary of the second comparison scenario can be seen in Table III.

Based on the two comparison scenarios, it can be concluded that BTM algorithm method to produce the topic modeling for reviewing the public hospital dataset extracted from mapping review service platform has acceptable accuracy and performance value. The result can be utilized by

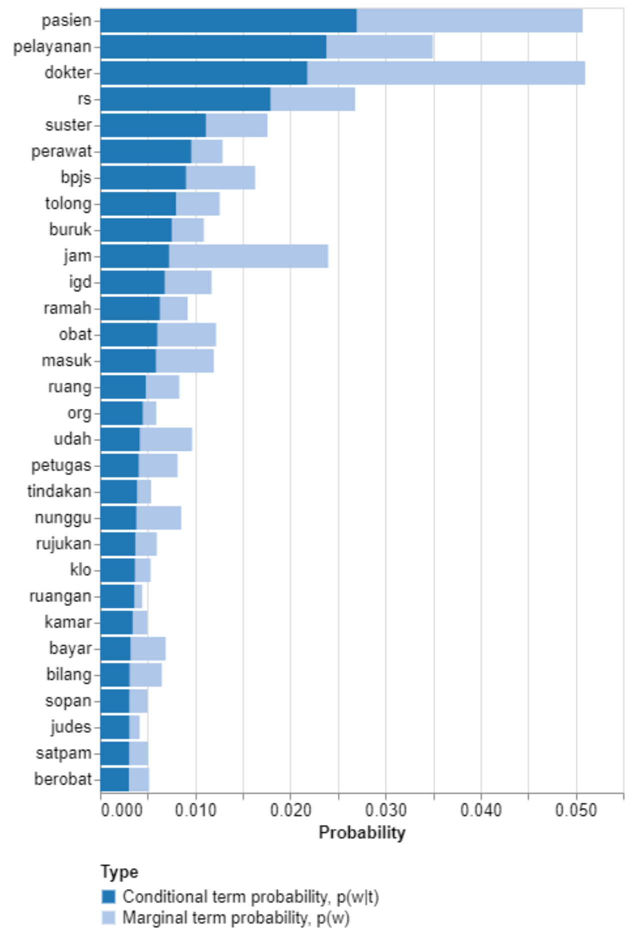

Figure 19. Topic Distribution for All Part-of-Speech Dataset

the health service provider to focus on the topics modeled in this research. Hence, the improvement of the service is oriented to the service user satisfaction. However, the topic creation using BTM algorithm by tagging noun-only is unable to increase the accuracy value significantly. This is because the NLTK part-of-speech tagger used in this study is in Indonesian which has a limited language resource. Hence, there is a need to have eminent natural language processing research that produces a much better part-of-speech tagger for the Indonesian language.

Drawing from the existing research [19], it can be inferred that Indonesian tag sets are currently not standardized and the number of Indonesian post tagger libraries are much less than the number of English post taggers. By using English, it is believed that the data can be appropriately labeled so that the BTM algorithm with an additional noun-only tagging approach can produce better performance and accuracy.

Another possible reason is that there are words that have not been properly labeled. This is because the review dataset contains a lot of informal or incomplete words, in contrast to the news datasets that use a standard language. It is expected to have more intensive data cleaning, although there is no robust technique that can solve the problem of document analysis written in informal language to the best of our knowledge. It is interesting to note, however, that the addition of noun-only selection could improve the overall performance, because the number of processed words is less than the one for all part-of-speech tags. Therefore, it increases efficiency by reducing the modeling time.

6. Conclusion

This research proposed to use the BTM algorithm method to produce the topic modeling that has satisfiable accuracy and performance value. The addition of noun-only tags to the algorithm has not been able to increase the accuracy value of the model due to several factors, such as the low quality of the latest part-of-speech tagger in the Indonesian language, and inappropriate word labeling influenced by informal language writing. However, the addition of noun-only tags to the algorithm can satisfy the performance requirement of the model by reducing the processing time. This research is expected to contribute to upgrade health services, especially in public hospitals, by focusing on the topic model obtained by the result.

For future work, the intensified data cleaning process for incomplete words is expected to influence performance improvement, although currently there is no robust technique that can solve the problem of document analysis written in informal language to the best of our knowledge. Moreover, upgrading the Indonesian part-of-speech tagger library is a must to have better topic modeling research in the Indonesian language. Another possible approach is to translate all reviews into English during data preparation. It is also suggested to have another dataset as a comparison when there will be another popular mapping review service platform.

References

[1] A. Andriani, “The relationship between health service quality and patient satisfaction in general clinic at bukittinggi health center [hubungan mutu pelayanan kesehatan dengan kepuasan pasien diruangan poli umum puskesmas bukittinggi],” *Jurnal Endurance: Kajian Ilmiah Problema Kesehatan*, vol. 2, no. 1, pp. 45–52, 2017.

[2] A. Bramantoro and T. Ishida, “User-centered qos in combining web services for interactive domain,” in *2009 Fifth International Conference on Semantics, Knowledge and Grid*. IEEE, 2009, pp. 41–48.

[3] I. G. A. Surati, G. A. Marhaeni, and N. W. Armini, “The satisfaction difference between the patient admitted to public and private hospitals in 2017 [perbedaan kepuasan pasien yang menggunakan jaminan kesehatan dirawat inap di ruang kebidanan rumah sakit pemerintah dengan swasta tahun 2017],” *Jurnal Ilmiah Kebidanan (The Journal Of Midwifery)*, vol. 6, no. 1, pp. 53–61, 2018.

[4] X. Yan, J. Guo, Y. Lan, and X. Cheng, “A biterm topic model for short texts,” in *Proceedings of the 22nd international conference on World Wide Web*, 2013, pp. 1445–1456.

[5] M. Lambovska and K. Yordanov, “Understanding ‘motivation-internal control’relations in municipalities,” *TEM Journal*, vol. 9, no. 2, p. 662, 2020.

[6] X. Cheng, X. Yan, Y. Lan, and J. Guo, “Btm: Topic modeling over short texts,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 12, pp. 2928–2941, 2014.

[7] A. Bramantoro, U. Schäfer, and T. Ishida, “Towards an integrated architecture for composite language services and multiple linguistic processing components,” in *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC’10)*, 2010, pp. 3506–3511.

[8] F. Martin and M. Johnson, “More efficient topic modelling through a noun only approach,” in *Proceedings of the Australasian Language Technology Association Workshop 2015*, 2015, pp. 111–115.

[9] D. M. Blei, A. Y. Ng, and M. I. Jordan, “Latent dirichlet allocation,” *Journal of machine Learning research*, vol. 3, no. Jan, pp. 993–1022, 2003.

[10] A. Daud, J. A. Khan, J. A. Nasir, R. A. Abbasi, N. R. Aljohani, and J. S. Alowibdi, “Latent dirichlet allocation and pos tags based method for external plagiarism detection: Lda and pos tags based plagiarism detection,” in *Scholarly Ethics and Publishing: Breakthroughs in Research and Practice*. IGI Global, 2019, pp. 319–336.

[11] I. Blekanov, N. Tarasov, and A. Maksimov, “Topic modeling of conflict ad hoc discussions in social networks,” in *Proceedings of the 3rd International Conference on Applications in Information Technology*, 2018, pp. 122–126.

[12] J. Schneider and M. Vlachos, “Topic modeling based on keywords and context,” in *Proceedings of the 2018 SIAM international conference on data mining*. SIAM, 2018, pp. 369–377.

[13] W. Li, Y. Feng, D. Li, and Z. Yu, “Micro-blog topic detection method based on btm topic model and k-means clustering algorithm,” *Automatic Control and Computer Sciences*, vol. 50, no. 4, pp. 271–277, 2016.

[14] K. Nimala and R. Jebakumar, “A robust user sentiment biterm topic mixture model based on user aggregation strategy to avoid data sparsity for short text,” *Journal of Medical Systems*, vol. 43, no. 4, pp. 1–13, 2019.

[15] W.-D. Zhu and W.-G. Zhou, “Short text topic discovery based on btm topic model,” in *6th International Conference on Management Science and Management Innovation (MSMI 2019)*. Atlantis Press, 2019, pp. 100–106.

[16] X. Li, A. Zhang, C. Li, L. Guo, W. Wang, and J. Ouyang, “Relational biterm topic model: Short-text topic modeling using word embeddings,” *The Computer Journal*, vol. 62, no. 3, pp. 359–372, 2019.

[17] X. He, H. Xu, J. Li, L. He, and L. Yu, “Fastbtm: Reducing the sampling time for biterm topic model,” *Knowledge-Based Systems*, vol. 132, pp. 11–20, 2017.

[18] A. E. Gelfand, “Gibbs sampling,” *Journal of the American statistical Association*, vol. 95, no. 452, pp. 1300–1304, 2000.

[19] K. M. Hana, S. Al Faraby, A. Bramantoro *et al.*, “Multi-label classification of indonesian hate speech on twitter using support vector machines,” in *2020 International Conference on Data Science and Its Applications (ICoDSA)*. IEEE, 2020, pp. 1–7.

Moh Makruf obtained his master’s degree in Computer Science at Universitas Budi Luhur Jakarta Indonesia in 2022. He is currently working as a software engineer at a private company in Indonesia. In 2012, he received a bachelor’s degree from Institut Sains dan Teknologi Nasional (ISTN), Indonesia with a specialization in informatics engineering. His research interest includes data science, machine learning, natural language processing, and artificial intelligence.

Arif Bramantoro is currently a senior assistant professor in the School of Computing and Informatics, Universiti Teknologi Brunei, Brunei Darussalam. Previously, he was an associate professor in the Information Systems Department, Faculty of Computing and Information Technology in Rabigh, King Abdulaziz University, Saudi Arabia. From 2012 to 2016, he was an assistant professor in the Information Systems Department, the College of Computer Sciences and Information, Al-Imam Mohammad Ibn Saud Islamic University, Riyadh, Saudi Arabia. He was an expert researcher at the Information Services Platform laboratory, National Institute of Information and Communication Technology, Japan, from 2011 to 2012. He received Ph.D. from the Department of Social Informatics, Kyoto University, Japan, in 2011. He holds master’s degree from the Faculty of Information Technology, Monash University, Melbourne, Australia in 2006. His bachelor’s degree was obtained from the Department of Informatics, Institute Technology of Bandung, Indonesia in 2001. His research interests include service systems, business process workflow, and business intelligence. He is the author of more than 50 articles. He was a recipient of the Research Excellence award in 2016 from the Deanship of Scientific Research, Al-Imam University, Saudi Arabia; and the Best Paper award from the IEEE International Conference on Cloud Computing in 2015.