CHAPTER 6
CONCLUSION

Through the table, Through this study, it can be concluded that the Latent Dirichlet Allocation Algorithm with the Gibbs Sampling method is proven to be able to do the scientific journal topic modeling based on the title. The use of the LDA algorithm is due to the need for special algorithms to handle this type of data that is too sparse. Journal data is declared sparse after passing the density calculation which reaches the number 0.197346369779%. The result of density calculation is obtained after going through the pre-processing process that changes 1047 documents to 4016 words.

Word results generated through the pre-processing process are reprocessed to be distributed to each topic using alpha and betta as 0.5 as hyperparameters. There are two types of distribution that are produced, namely the distribution of words and topic documents. The number of topics used is 3 topics. Determining the number of topics is done at the testing stage by comparing the perplexity.

In the sampling stage, it is done using the Gibbs Sampling method. The sampling process is carried out repeatedly as much as iteration. Iteration is a specified parameter. The process of determining the number of iterations also uses perplexity by comparing the perplexity of each candidate. Based on the results of the comparison, iteration 500 is considered the most optimal in this study.

The suggestion for further research is to use a larger amount of document data so that the distribution of data is increasingly apparent. it can be seen that the number of repetitions is very influential on the calculation of probability. The following is a graph of each candidate repetition to find the number of repetitions that reach minimum points.