Assignment: Text analytics assignment 1
Name: Shashi Lokesh
Registration number:2100459

Topic Tuning Process: The researchers of [1] have applied the LDA method by implying Topic-Word Matrix, Document-Word Matrix, Symmetric Kullback-Leibler divergence to find the number of topics. They have found 100 topics based on the tuning of the number of documents. The researchers of [2] have proposed the model to find the number of topics in a document using the LDA model and with the implication of the Document-Term Matrix through the t-SNE mechanism. In this research, they have shown that the number of topics in a document can be varied with the variation of the Perplexity of the document. **Text Summarization Process**: The researchers of [3] have applied the LDA model to summarise the text document and uses internal dataset which consist of 398 articles from public blog with the implication of TF-IDF and text clustering. The researchers of [4] have selected government agency data for text summarization and conducted the research by applying the LDA model. In this research, for the modelling of topics by finding the number of topics, they have used a part-of-speech tagger along with the LDA. **Text Vectorization Process**: The researchers of [5] have selected 1824 SCOPUS papers and applied the LDA model using TF-IDF and N-Gram analysis to find the number of topics through text analytics and text summarization of both the title and abstract of articles published until 2019. The researchers of [6] have applied the LDA model for the summarization and modelling of Twitter text data using TF-IDF. In this research, they have collected tweet data with four topics namely Economic, Military, Sports, Technology and applied the mechanisms for topic finding. The researchers of [7] have proposed the topic modelling technique using the LDA method. In this research, they have selected the tweet data for Central Java Gubernatorial Election and applied the Naïve Bayes & Support Vector method followed by Text Vectorixzation through the bag of words.

References

[1]. R. Arun, V. Suresh, C. E. V. Madhavan and M. N. N. Murthy, "On Finding the Natural Number of Topics with Latent Dirichlet Allocation: Some Observations," Advances in Knowledge Discovery and Data Mining, pp. 391-402, 2010.

[2]. N. Horn and F. G. a. R. Buchkremer, "Latent Dirichlet Allocation and tDistributed Stochastic Neighbor Embedding Enhance Scientific Reading Comprehension of Articles Related to Enterprise Architecture," MDPI, pp. 179195, 2021.

[3]. E. Y. H. Mail, F. F. Mail, K. H. Mail and I. N. D. M. a. A. Azhari, "Automatic Text Summarization Using Latent Drichlet Allocation (LDA) for Document Clustering," International Journal in Advances in Intelligent Informatics, vol. 1, no. 3, pp. 1-7, 2015.

[4]. D. Wibawa and R. E. Listanto, "Complaint Data Text Analysis Concerning the Apps Provided by Government Agency using Inference LDA," 8th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI), pp. 1-5, 2021.

[5]. M. F. Hilmi, Y. Mustapha and M. T. C. Omar, "Innovation in an Emerging Market: A Bibliometric and Latent Dirichlet Allocation Based Topic Modeling Study," International Conference on Decision Aid Sciences and Application (DASA), pp. 1-5, 2020.

[6]. E. S. Negara, D. Triadi and R. Andryani, "Topic Modelling Twitter Data with Latent Dirichlet Allocation Method," International Conference on Electrical Engineering and Computer Science (ICECOS), pp. 1-4, 2019.
[7]. G. R. G. Wisnu, Ahmadi, A. R. Muttaqi, A. B. Santoso, P. K. Putra and I. Budi, "Sentiment Analysis and Topic Modelling of 2018 Central Java Gubernatorial Election using Twitter Data," International Workshop on Big Data and Information Security (IWBIS), pp. 1-5, 2020.