

DAFTAR PUSTAKA

- Ader, H., Mellenbergh, G. dan Hand, D., 2011. *Advising on Research Methods: A Consultant's Companion*, Vol. 38.
- Alasadi, S., 2017, Review of data preprocessing techniques in data mining, *Journal of Engineering and Applied Sciences* **12**: 4102–4107.
- Blei, D., Carin, L. dan Dunson, D., 2010, Probabilistic topic models: A focus on graphical model design and applications to document and image analysis, *IEEE signal processing magazine* **27**: 55–65.
- Blei, D. M., Ng, A. Y. dan Jordan, M. I., 2003, Latent dirichlet allocation, *J. Mach. Learn. Res.* **3**: 993–1022. <http://dl.acm.org/citation.cfm?id=944919.944937>
- Boeing, G. dan Waddell, P., 2017, New insights into rental housing markets across the united states: Web scraping and analyzing craigslist rental listings, *Journal of Planning Education and Research* **37**.
- Bunjamin, H. dan Sulistiani, L., (2017). Automatic topic clustering using latent dirichlet allocation with skip-gram model on final project abstracts, *2017 21st International Computer Science and Engineering Conference (ICSEC)*, pp. 1–5.
- Chen, Z., Zhang, C., Zhao, Z., Yao, C. dan Cai, D., 2018, Question retrieval for community-based question answering via heterogeneous social influential network, *Neurocomputing* **285**: 117 – 124. <http://www.sciencedirect.com/science/article/pii/S0925231218300523>
- Cox, D. R., 2006. *Principles of Statistical Inference*, Cambridge University Press.
- Dwivedi, S. K. dan Singh, V., 2013, Research and reviews in question answering system, *Procedia Technology* **10**: 417 – 424. First International Conference on Computational Intelligence: Modeling Techniques and Applications (CIMTA) 2013. <http://www.sciencedirect.com/science/article/pii/S2212017313005409>
- Fan, W., Wallace, L., Rich, S. dan Zhang, Z., 2006, Tapping the power of text mining, *Communications of the ACM* **49**: 76–82.

- Figueroa, A., 2017, Automatically generating effective search queries directly from community question-answering questions for finding related questions, *Expert Systems with Applications* **77**: 11 – 19. <http://www.sciencedirect.com/science/article/pii/S0957417417300556>
- Gimpel, K., (2006). Modeling topics.
- Greene, D., O’Callaghan, D. dan Cunningham, P., (2014). How many topics? stability analysis for topic models.
- Griffiths, T. L. dan Steyvers, M., 2004, Finding scientific topics, *Proceedings of the National Academy of Sciences* **101**(Suppl. 1): 5228–5235.
- Grün, B. dan Hornik, K., 2011, topicmodels: An r package for fitting topic models, *Journal of Statistical Software, Articles* **40**(13): 1–30. <https://www.jstatsoft.org/v040/i13>
- Hakim, A. N., Mahendra, R., Adriani, M. dan Ekakristi, A. S., (2017). Corpus development for indonesian consumer-health question answering system, *2017 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, pp. 222–227.
- Hong, W., Zheng, X., Qi, J., Wang, W. dan Weng, Y., (2018). Project rank: An internet topic evaluation model based on latent dirichlet allocation, *2018 13th International Conference on Computer Science Education (ICCSE)*, pp. 1–4.
- Kaur, P. C., Ghorpade, T. dan Mane, V., (2017). Extraction of unigram and bigram topic list by using latent dirichlet markov allocation and sentiment classification, *2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS)*, pp. 2332–2339.
- Kotu, V. dan Deshpande, B., (2019). Chapter 9 - text mining, in V. Kotu dan B. Deshpande (eds), *Data Science (Second Edition)*, second edition edn, Morgan Kaufmann, pp. 281 – 305. <http://www.sciencedirect.com/science/article/pii/B9780128147610000095>
- Kurdi, M., 2016. *Natural Language Processing and Computational Linguistics 1: Speech, Morphology and Syntax*.

- Laoh, E., Surjandari, I. dan Febirautami, L. R., (2018). Indonesians' song lyrics topic modelling using latent dirichlet allocation, *2018 5th International Conference on Information Science and Control Engineering (ICISCE)*, pp. 270–274.
- Lee, D. D. dan Seung, H. S., (2001). Algorithms for non-negative matrix factorization, *Advances in neural information processing systems*, pp. 556–562.
- Lin, J., 1991, Divergence measures based on the shannon entropy, *IEEE Transactions on Information Theory* **37**(1): 145–151.
- Michie, D., Spiegelhalter, D. J., Taylor, C. C. dan Campbell, J. (eds), 1994. *Machine Learning, Neural and Statistical Classification*, Ellis Horwood, Upper Saddle River, NJ, USA.
- Mitchell, T. M., (2005). Chapter 3 generative and discriminative classifiers : Naive bayes and logistic regression machine learning.
- Ng, A. Y. dan Jordan, M. I., (2002). On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes, in T. G. Dietterich, S. Becker dan Z. Ghahramani (eds), *Advances in Neural Information Processing Systems 14*, MIT Press, pp. 841–848. <http://papers.nips.cc/paper/2000-on-discriminative-vs-generative-classifiers-a-comparison-of-logistic-regression-and-naive-bayes.pdf>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. dan Duchesnay, E., 2011, Scikit-learn: Machine learning in Python, *Journal of Machine Learning Research* **12**: 2825–2830.
- Qaiser, S. dan Ali, R., 2018, Text mining: Use of tf-idf to examine the relevance of words to documents, *International Journal of Computer Applications* **181**.
- Rajaraman, A. dan Ullman, J. D., 2011. *Mining of Massive Datasets*, Cambridge University Press, New York, NY, USA.
- Řehůřek, R. dan Sojka, P., (2010). Software Framework for Topic Modelling with Large Corpora, *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, ELRA, Valletta, Malta, pp. 45–50. <http://is.muni.cz/publication/884893/en>.
- Richardson, L., 2007, Beautiful soup documentation, *April* .

- Samuel, H., Kim, M., Prabhakar, S., Jabbar, M. S. M. dan Zalane, O., (2017). Community question retrieval in health forums, *2017 IEEE EMBS International Conference on Biomedical Health Informatics (BHI)*, pp. 77–80.
- Sievert, C. dan Shirley, K., (2014). Ldavis: A method for visualizing and interpreting topics.
- Sorokin, D. dan Gurevych, I., (2018). Modeling semantics with gated graph neural networks for knowledge base question answering, *Proceedings of the 27th International Conference on Computational Linguistics*, Association for Computational Linguistics, Santa Fe, New Mexico, USA, pp. 3306–3317. <https://www.aclweb.org/anthology/C18-1280>
- Srba, I. dan Bielikova, M., 2016, A comprehensive survey and classification of approaches for community question answering, *ACM Transactions on the Web* **10**.
- Stevens, K., Kegelmeyer, P., Andrzejewski, D. dan Buttler, D., (2012). Exploring topic coherence over many models and many topics, *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, Association for Computational Linguistics, Jeju Island, Korea, pp. 952–961. <https://www.aclweb.org/anthology/D12-1087>
- Voorhees, E. dan Harman, D., 2000, Proceedings of the eighth text retrieval conference (trec-8).
- Wang, Y., Mehrabi, S., Mojarad, M. R., Li, D. dan Liu, H., (2015). Retrieval of semantically similar healthcare questions in healthcare forums, *2015 International Conference on Healthcare Informatics*, pp. 517–518.
- Weiss, S. M., Indurkha, N., Zhang, T. dan Damerau, F., 2010. *Text Mining: Predictive Methods for Analyzing Unstructured Information*, 1st edn, Springer Publishing Company, Incorporated.
- Yang Bao, Zhu Qibing dan Huang Min, (2012). Image identification of glass defects based on non-negative matrix factorization and sparse representation classification, *2012 24th Chinese Control and Decision Conference (CCDC)*, pp. 3225–3229.
- Yeh, J., Lee, C., Tan, Y. dan Yu, L., (2014). Topic model allocation of conversational dialogue records by latent dirichlet allocation, *Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2014 Asia-Pacific*, pp. 1–4.

- Yildirim, I., 2012, Bayesian inference: Gibbs sampling, *Technical Note, University of Rochester*.
- Zhang, K., Wu, W., Wu, H., Li, Z. dan Zhou, M., (2014). Question retrieval with high quality answers in community question answering, *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management, CIKM '14*, ACM, New York, NY, USA, pp. 371–380. <http://doi.acm.org/10.1145/2661829.2661908>
- Zhou, G., Xie, Z., He, T., Zhao, J. dan Hu, X. T., 2016, Learning the multilingual translation representations for question retrieval in community question answering via non-negative matrix factorization, *IEEE/ACM Transactions on Audio, Speech, and Language Processing* **24**(7): 1305–1314.
- Zou, Q., c. Sun, Yuan, C. dan Sun, Y., (2019). Non-negative matrix factorization speech enhancement method based on constraints of temporal continuity, *2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*, pp. 542–546.