

DAFTAR PUSTAKA

- [1] F. Yuan, L. Yang, and G. Yu, "A new density-based method for reducing the amount of training data in K-NN text classification," *Proc. Sixth Int. Conf. Mach. Learn. Cybern. ICMLC 2007*, vol. 6, no. August, pp. 3372–3376, 2007.
- [2] K. Yang *et al.*, "A new samples selecting method based on K nearest neighbors," *2017 IEEE Int. Conf. Big Data Smart Comput. BigComp 2017*, pp. 457–462, 2017.
- [3] H. Yu, J. Yang, and J. Han, "Classifying large data sets using SVMs with hierarchical clusters," *9th ACM SIGKDD Int. Conf. Knowl. Discov. data Min.*, pp. 306–314, 2003.
- [4] Subdirektorat Statistik Transportasi, *Statistik Transportasi 2010*. Badan Pusat Statistik, 2011.
- [5] Subdirektorat Statistik Transportasi, *Statistik Transportasi 2011*. Badan Pusat Statistik, 2012.
- [6] Subdirektorat Statistik Transportasi, *Statistik Transportasi 2012*. Badan Pusat Statistik, 2013.
- [7] Subdirektorat Statistik Transportasi, *Statistik Transportasi 2013*. Badan Pusat Statistik, 2014.
- [8] Subdirektorat Statistik Transportasi, *Statistik Transportasi Darat 2014*. Badan Pusat Statistik, 2015.
- [9] Subdirektorat Statistik Transportasi, *Statistik Transportasi Darat 2015*. Badan Pusat Statistik, 2016.
- [10] T. Sakaki, M. Okazaki, and Y. Matsuo, "Earthquake shakes Twitter users," in *Proceedings of the 19th international conference on World wide web - WWW '10*, 2010, p. 851.
- [11] E. D'Andrea, P. Ducange, B. Lazzarini, and F. Marcelloni, "Real-Time Detection of Traffic From Twitter Stream Analysis," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 4, pp. 2269–2283, Aug. 2015.
- [12] T. Sakaki, Y. Matsuo, T. Yanagihara, N. P. Chandrasiri, and K. Nawa, "Real-time event extraction for driving information from social sensors," in *Proceedings - 2012 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems, CYBER 2012*, 2012, pp. 221–226.
- [13] Y. Gu, Z. (Sean) Qian, and F. Chen, "From Twitter to detector: Real-time traffic incident detection using social media data," *Transp. Res. Part C Emerg. Technol.*, vol. 67, pp. 321–342, Jun. 2016.
- [14] C. Gutierrez, P. Figuerias, P. Oliveira, R. Costa, and R. Jardim-Goncalves, "Twitter mining for traffic events detection," in *Proceedings of the 2015 Science and Information Conference, SAI 2015*, 2015, pp. 371–378.
- [15] A. Wibisono *et al.*, "Traffic Intelligent System Architecture Based on Social Media Information," *Adv. Comput. Sci. Inf. Syst. (ICACSIS), 2012 Int. Conf.*, pp. 25–30, Dec. 2012.
- [16] S. F. Rodiyansyah and E. Winarko, "Klasifikasi Posting Twitter Kemacetan

- Lalu Lintas Kota Bandung Menggunakan Naive Bayesian Classification,” *IJCCS-Indonesian J. Comput. Cybern. Syst.*, vol. 7, no. 1, pp. 13–22, 2013.
- [17] D. A. Kurniawan, “Analisis Data Jejaring Sosial Twitter Untuk Pemetaan Kondisi Kemacetan Jalan Di Provinsi DIY Dengan Metode Text Mining,” Universitas Gadjah Mada, Yogyakarta, Indonesia, 2016.
- [18] T. Joachims, “Text categorization with support vector machines: Learning with many relevant features,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 1398, no. LS-8 Report 23, pp. 137–142, 1998.
- [19] C.-W. Hsu, C.-C. Chang, and C.-J. Lin, “A Practical Guide to Support Vector Classification,” 2003.
- [20] D. A. Kurniawan, S. Wibirama, and N. A. Setiawan, “Real-time traffic classification with Twitter data mining,” in *2016 8th International Conference on Information Technology and Electrical Engineering (ICITEE)*, 2016, pp. 1–5.
- [21] R. K. Anas, “APLIKASI PEMANTAUAN LALU LINTAS YOGYAKARTA MENGGUNAKAN GET SOURCE TWITTER DAN GOOGLE MAPS API BERBASIS WEB,” *SEMNASTEKNOMEDIA ONLINE*, vol. 4, no. 1, pp. 2–11, 2016.
- [22] T. Zhang, R. Ramakrishnan, and M. Livny, “BIRCH: An Efficient Data Clustering Databases Method for Very Large,” *ACM SIGMOD Int. Conf. Manag. Data*, vol. 1, pp. 103–114, 1996.
- [23] A. M. Ben Omran and M. J. Ab Aziz, “Automatic essay grading system for short answers in English language,” *J. Comput. Sci.*, vol. 9, no. 10, pp. 1369–1382, Oct. 2013.
- [24] W. Gomaa and A. Fahmy, “Short Answer Grading Using String Similarity And Corpus-Based Similarity,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 3, no. 11, pp. 115–121, 2012.
- [25] W. Lu, X. Du, M. Hadjieleftheriou, and B. C. Ooi, “Efficiently Supporting Edit Distance Based String Similarity Search Using B+-Trees,” *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 12, pp. 2983–2996, Dec. 2014.
- [26] D. Deng, G. Li, J. Feng, and W. S. Li, “Top-k string similarity search with edit-distance constraints,” in *Proceedings - International Conference on Data Engineering*, 2013, pp. 925–936.
- [27] Y. Zheng, T. Deng, and X. Wang, “Light-weight top-k string similarity search with a two-level inverted index,” 2016, pp. 718–723.
- [28] S. Niwattanakul, J. Singthongchai, E. Naenudorn, and S. Wanapu, “Using of Jaccard Coefficient for Keywords Similarity,” *Int. MultiConference Eng. Comput. Sci.*, vol. I, pp. 380–384, 2013.
- [29] V. Thada and V. Jaglan, “Comparison of Jaccard, Dice, Cosine Similarity Coefficient To Find Best Fitness Value for Web Retrieved Documents Using Genetic Algorithm,” *Int. J. Innov. Eng. Technol.*, vol. 2, no. 4, pp. 202–205, 2013.
- [30] W. H. Gomaa and A. A. Fahmy, “A Survey of Text Similarity Approaches,” *Int. J. Comput. Appl.*, vol. 68, no. 13, pp. 13–18, Apr. 2013.
- [31] R. Manurung, B. Distiawan, and D. D. Putra, “Developing an online

- Indonesian corpora repository,” in *PACLIC*, 2010, pp. 243–249.
- [32] S. Montalvo, E. G. Pardo, R. Martinez, and V. Fresno, “Automatic cognate identification based on a fuzzy combination of string similarity measures,” in *IEEE International Conference on Fuzzy Systems*, 2012, pp. 1–8.
- [33] T. B. Adji, Z. Abidin, and H. A. Nugroho, “System of negative Indonesian website detection using TF-IDF and Vector Space Model,” in *Proceedings of 2014 International Conference on Electrical Engineering and Computer Science, ICEECS 2014*, 2014, pp. 174–178.
- [34] N. Monarizqa, L. E. Nugroho, and B. S. Hantono, “Penerapan Analisis Sentimen Pada Twitter Berbahasa Indonesia Sebagai Pemberi Rating,” Universitas Gadjah Mada, Yogyakarta, Indonesia, 2014.
- [35] “Tweepy.” [Online]. Available: <http://www.tweepy.org/>. [Accessed: 29-Apr-2016].
- [36] J. dan M. K. Han, *Data Mining: Concept and Techniques*, 3rd ed. Burlington, MA: Elsevier, 2001.
- [37] C. de la Torre and M. Martín-Bautista, “Text mining: intermediate forms on knowledge representation,” *EUSFLAT Conf.*, pp. 1082–1087, 2005.
- [38] S. Bird, E. Klein, and E. Loper, *Natural language processing with Python*, 1st ed. Beijing ; Cambridge [Mass.]: O’Reilly, 2009.
- [39] C. D. Manning, P. Raghavan, and H. Schütze, “An Introduction to Information Retrieval,” *Online*, vol. 1, no. c, p. 569, 2009.
- [40] C. Rong, W. Lu, X. Wang, X. Du, Y. Chen, and A. K. H. Tung, “Efficient and scalable processing of string similarity join,” *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 10, pp. 2217–2230, Oct. 2013.
- [41] H. Hu, K. Zheng, X. Wang, and A. Zhou, “GFilter: A General Gram Filter for String Similarity Search,” *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 4, pp. 1005–1018, Apr. 2015.
- [42] E. H. Ko and D. Klabjan, “Semantic properties of customer sentiment in tweets,” in *Proceedings - 2014 IEEE 28th International Conference on Advanced Information Networking and Applications Workshops, IEEE WAINA 2014*, 2014, pp. 657–663.
- [43] L. R. . Dice, “Measures of the Amount of Ecologic Association Between Species,” *Ecology*, vol. 26, no. 3, pp. 297–302, 1945.
- [44] R. Garreta and G. Moncecchi, *Learning scikit-learn: machine learning in Python : experience the benefits of machine learning techniques by applying them to real-world problems using Python and the open source scikit-learn library*. 2013.