

DAFTAR PUSTAKA

- Adiwijawa, I., 2006, *Text Mining dan Knowledge Discovery*, EMC Corporation, Hopkinton.
- Adriyani, N.M., Santiyasa, I.W., & Muliantara, A., 2012, Implementasi Algoritma Levenshtein Distance dan Metode Empiris untuk Menampilkan Saran Perbaikan Kesalahan Pengetikan Dokumen Berbahasa Indonesia, *Skripsi*, Universitas Udayana, Bali.
- Agbehadji, I., Yang, H., Fong, S., & Millham, R., 2018, The Comparative Analysis of Smith-Waterman Algorithm with Jaro-Winkler Algorithm for the Detection of Duplicate Health-Related Records, *2018 International Conference on Advances in Big Data, Computing and Data Communication Systems (icABCD)*, Durban, 6-7 Agustus 2018.
- Ahmed, B., 2015, Lexical normalisation of Twitter Data, *Science and Information Conference (SAI)*, London, 28-30 Juli 2015.
- Akinwale, A. & Niewiadomski, A., 2015, Efficient Similarity Measures for Texts Matching, *Journal of Applied Computer Science*, 1, 23, 7 – 28.
- Antinasari, P., Perdana, R.S., & Fauzi, M.A., 2017, Analisis Sentimen Tentang Opini Film Pada Dokumen Twitter Berbahasa Indonesia Menggunakan Naive Bayes Dengan Perbaikan Kata Tidak Baku, *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 12, 1, 1733 – 1741.
- Bakr, H. A., Shaalan, K., & Ziedan, I., 2008, A Hybrid Approach for Converting Written Egyptian Colloquial Dialect into Diacritized Arabic, *The 6th International Conference on Informatics and Systems*, Cairo.
- Chandrasekar, P., & Qian, K., 2016, The Impact of *Data preprocessing* on the Performance of a Naive Bayes Classifier, *2016 IEEE 40th Annual Computer Software and Applications Conference (COMPSAC)*, Atlanta, 10-14 Juni 2016.

- Cook, P., & Stevenson, S., 2009, An Unsupervised Model for Text Message Normalization, *Proceedings of The Workshop on Computational Approaches to Linguistic Creativity*, Boulder.
- Duque, T., 2020, Text Normalization, <https://medium.com/swlh/text-normalization-7ecc8e084e31>, diakses tanggal 7 Juli 2020.
- Feldman, R., & Sanger, J., 2007, *Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*, New York, Cambridge University Press.
- Fawcett, T., 2006, An introduction to ROC analysis, *Pattern recognition letters*, 27, 8, 861-874.
- Han, J., Pei, J., & Kamber, M., 2006, *Data Mining, Southeast Asia Edition: Concepts and Techniques*, Burlington, Morgan Kaufmann.
- Imbar, R.V., Adelia, Ayub, M., & Rehatta, A., 2014, Implementasi Cosine Similarity dan Algoritma Smith-Waterman untuk Mendeteksi Kemiripan Teks, *Jurnal Informatika*, 1, 10, 31-42.
- Kadiri, J.A., & Adetoro, N.A., 2012, Information Explosion and the Challenges of Information and Communication Technology Utilization in Nigerian Libraries and Information Centres, *Ozean Journal of Social Sciences*, 1, 5, 21-30.
- Kaplan, R.M., 2005, *A Method for Tokenizing Text*, Individual Authors, Complexity and Education: Inquiries into Words, *Constraints and Contexts*, Stanford, CSLI publications.
- Kobus, C., Yvon, F., & Damnati, G., 2008, Normalizing SMS: Are Two Metaphors Better than One?, *Proceedings of the 22nd International Conference on, Computational Linguistics-Volume 1*, Manchester.
- Kusumawardani, R., PradinaPriansya, S., & Atletiko, F.J., 2018, Context-sensitive normalization of social media text in bahasa Indonesia based on neural word embeddings, *Procedia Computer Science*, 1, 144, 105-117.

- Lailiyah, M., 2017, Sentiment Analysis Menggunakan Rule Based Method pada Data Pengaduan Publik Berbasis Lexical Resources, *Tesis*, Fakultas Teknologi Elektro, Institut Teknologi Sepuluh Nopember, Surabaya.
- Lisangan, E.A., 2015, Implementasi n-gram Technique dalam Deteksi Plagiarisme pada Tugas Mahasiswa, *Skripsi*, Fakultas Teknologi Informasi, Universitas Atma Jaya, Makassar.
- Manning, C. D., Schütze, H., & Raghavan, P., 2008, *Introduction to Information Retrieval* (Vol. 39), Cambridge University Press, Cambridge.
- Mulyanto, A., 2010, Analisis Edit Distance Menggunakan Algoritma Dynamic Programming, *Skripsi*, Fakultas Teknik, Universitas Negeri Gorontalo, Gorontalo.
- Mustaqhiri, M., 2011, Peringkasan Teks Otomatis Berita Berbahasa Indonesia Menggunakan Metode Maximum Marginal Relevance, *Skripsi*, Jurusan Teknik Informatika, Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim, Malang.
- Nugraha, I., & Rizqullah, R., 2019, Normalisasi Kata Tidak Baku yang Tidak Disingkat dengan Jarak Perubahan, *Jurnal Nasional Teknik Elektro dan Teknologi Informasi (JNTETI)*, 3, 8, 218-224.
- Poolsukkho, S., & Kongkachandra, R., 2018, Text Normalization on Thai Twitter Messages using IPA Similarity Algorithm, *2018 International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP)*, Pattaya.
- Raju, S.K. & Murty, M.R., 2018, Support Vector Machine with K-fold Cross Validation Model for Software Fault Prediction, *International Journal of Pure and Applied Mathematics*, 20, 118, 321 – 334.
- Rochmawati, Y., & Kusumaningrum, R., 2016, Studi Perbandingan Algoritma Pencarian String dalam Metode Approximate String Matching untuk

Identifikasi Kesalahan Pengetikan Teks, *Jurnal Buana Informatika*, 2, 7, 125-134.

Salsabila, N.A., Winatmoko, Y.A., Septiandri, A.A., & Jamal, A., 2018, Colloquial Indonesian Lexicon, *International Conference on Asian Language Processing (IALP)*, Bandung, 15-17 November 2018.

Saputra, Y., 2017, Perbandingan Metode Cosine Similarity Dan Algoritma Smith-Waterman Dalam Pendeteksian Kemiripan Text, *Skripsi*, Fakultas Ilmu Komputer, Universitas Sriwijaya, Palembang.

Saragih, T.S., 2017, Normalisasi Teks pada Teks Twitter Berbahasa Indonesia Menggunakan Algoritme Jarak String pada R, *Skripsi*, Fakultas Matematika dan Ilmu Pengetahuan Alam, Institut Pertanian Bogor, Bogor.

Sproat, R., Black, A.W., Chen, S., Kumar, S., Ostendorf, M., & Richards, C., 2001, Normalization of Non-Standard Words, *Computer Speech and Language*, 3, 15, 287-333.

Taheri, S., & Mammadov, M., 2013, Learning The Naive Bayes Classifier With Optimization Models, *International Journal of Applied Mathematics and Computer Science*, 31 Desember 2013, 787-795.

Thalib, F., & Kusumawati, R., 2014, Pembuatan Program Aplikasi untuk Pendeteksian Kemiripan Dokumen Teks dengan Algoritma Smith-Waterman, *Laboratorium Sistem Komputer Dan Sistem Informasi*. Universitas Gunadarma, Depok.

Torunoğlu, D., Çakirman, E., Ganiz, M., Akyokuş, S., & Gürbüz, M.Z., 2011, Analysis of Preprocessing Methods on Classification of Turkish Texts, *2011 International Symposium on Innovations in Intelligent Systems and Applications*, Istanbul, 15-18 Juni 2011.

Winkler, W.E., 1990, String Comparator Metrics and Enhanced Decision Rules in the Fellegi-Sunter Model of Record Linkage, *Proceedings of the Section on Survey Research*, Washington, D.C.

Van der Loo, M.P., 2014, The Stringdist Package for Approximate String Matching, *The R Journal*, 1, 6, 111–122.