DAFTAR PUSTAKA

Bilenko, M., & Mooney, R. J. (2002). Learning to Combine Trained Distance Metrics for Duplicate Detection in Databases. *Technical Report AI-02-296, Artificial Intelligence Lab, University of Texas at Austin*, (February), 1–19.

Hanrong, L., Xin, C., Xuhui, L., & Feng, Z. (2016). Duplicate data detection using GNN. *IEEE International Conference on Cloud Computing and Big Data Analysis(ICCCBDA)*, 167–170. https://doi.org/10.1109/ICCCBDA.2016.7529552

Hernandéz, M., & Stolfo, S. (1995). The Merge / Purge Problem for Large Databases. *SIGMOD '95 Proceedings of the 1995 ACM SIGMOD International Conference on Management of Data*, 127–138. https://doi.org/10.1145/568271.223807

Hidayatullah, A. F., Prasetyo, A. D., & Sari, D. P. (2014). Analisis Kualitas Data dan Klasifikasi Data Pasien Kanker. *Seminar Nasional Informatika Medis*, 38–47.

Imbar, R. V., Ayub, M., & Rehatta, A. (2014). Implementasi Cosine Similarity dan Algoritma Smith-Waterman untuk Mendeteksi Kemiripan Teks. *Jurnal Informatika*, *10*, 31–42.

Lisangan, E. A. (2013). Implementasi n-Gram Technique Dalam Deteksi Plagiarism Pada Tugas Mahasiswa. *TEMATIKA, Journal of Informatics and Information Systems*, *1*(2), 24–30. Retrieved from https://tematika.uajm.ac.id/index.php/tematika/article/view/10

Mualifa, R. (2016). Rancang Bangun Sistem Data Cleaning untuk Master Data Konsumen di PT XYZ Dengan Menerapkan Metode Sorted Neighbourhood dan Metode N-gram.

Priya, M., Kalpana, R., & Srisupriya, T. (2017). Hybrid Optimization Algorithm Using N Gram Based Edit Distance, 216–221.

Priyanka, M., & Baby, A. (2017). A Survey on Various Duplicate Detection Methods, *8*(1), 7–9.

Samiei, A., & Naumann, F. (2016). Cluster-based Sorted Neighborhood for Efficient Duplicate Detection. https://doi.org/10.1109/ICDMW.2016.55

Santra, A. K., & Christy, C. J. (2012). Genetic Algorithm and Confusion Matrix for Document Clustering, *9*(1), 322–328.

Sessa, J., & Syed, D. (2016). Techniques to Deal with Missing Data. *IEEE Access*.

Sureka, A., & Jalote, P. (2010). Detecting duplicate bug report using character N-gram-based features. *Proceedings - Asia-Pacific Software Engineering Conference, APSEC*, 366–374. https://doi.org/10.1109/APSEC.2010.49

Tian, Z., Lu, H., Ji, W., Zhou, A., & Tian, Z. (2001). An n-gram-based approach for detecting approximately duplicate database records. *International Journal on Digital Libraries*, *3*(4), 325–331. https://doi.org/10.1007/s007990100044

V, D. G., & Kumaresan, P. . (2016). A Survey On Duplicate Record Detection In Real World Data. *3rd International Conference on Advanced Computing and Communication Systems (ICACCS -2016)*.

Arfa Skandar, M. R. (2015). An Efficient Duplication Record Detection Algorithm for Data Cleansing. *International Journal of Computer Applications, 127*