

BAB V

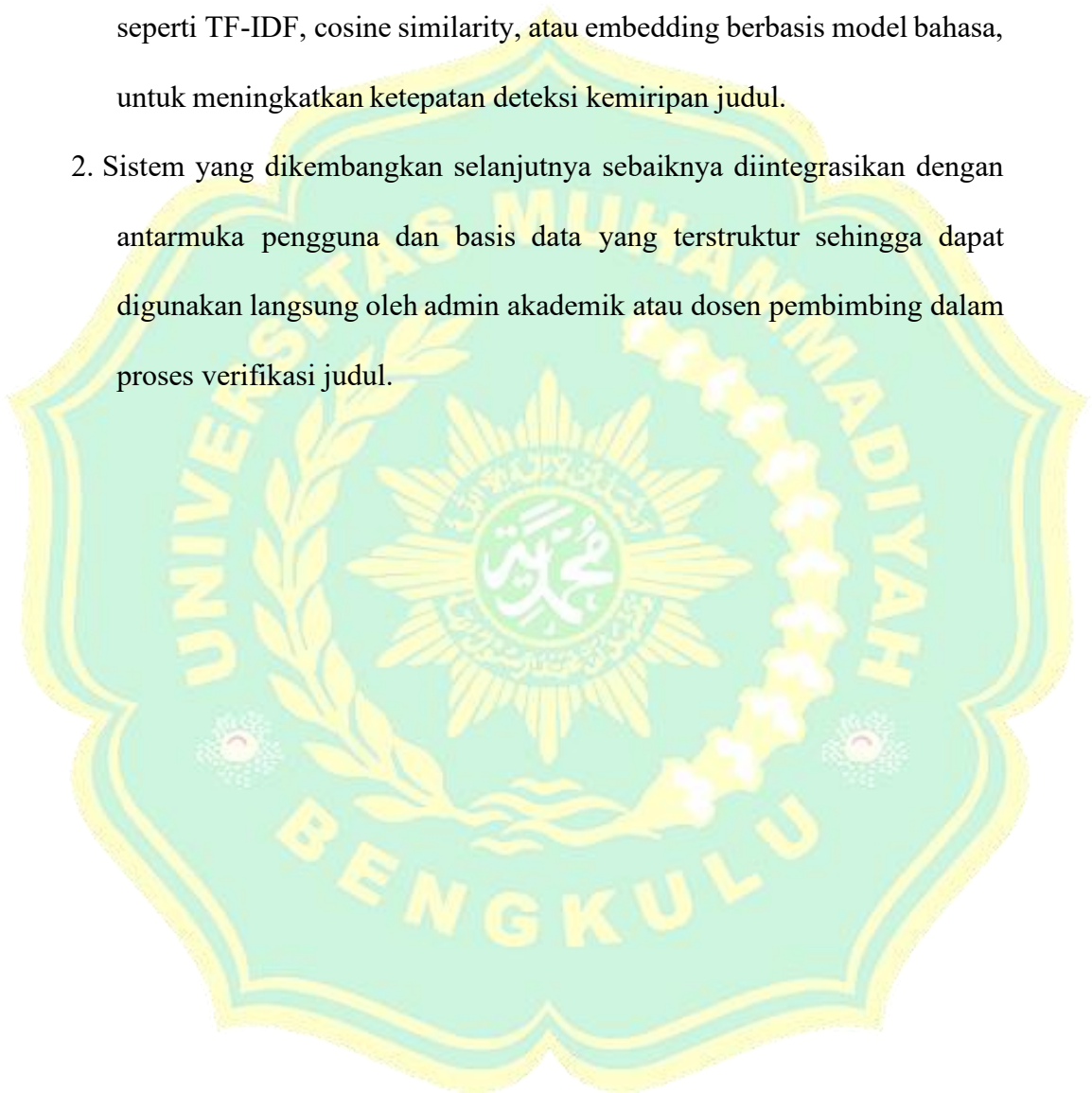
KESIMPULAN DAN SARAN

5.1 Kesimpulan

1. Algoritma Jaro-Winkler dan Levenshtein sama-sama dapat digunakan untuk mendeteksi kemiripan judul skripsi, namun pada penelitian ini Jaro-Winkler menunjukkan kinerja yang lebih baik. Hal ini dibuktikan oleh nilai F1-score tertinggi sebesar 1,000 pada threshold 0,85 dengan hasil klasifikasi yang sepenuhnya benar (TP=25, FP=0, FN=0, TN=25).
2. Variasi threshold similarity terbukti memengaruhi hasil klasifikasi secara langsung. Threshold yang terlalu rendah cenderung meningkatkan false positive, sedangkan threshold yang terlalu tinggi menurunkan recall karena sejumlah pasangan judul yang sebenarnya mirip tidak lagi terdeteksi.
3. Threshold optimal pada algoritma Levenshtein dalam penelitian ini berada pada nilai 0,70 dengan F1-score 0,837. Pada nilai threshold tersebut, sistem masih mampu menjaga keseimbangan antara precision dan recall, meskipun performanya belum melampaui Jaro-Winkler.
4. Hasil uji Wilcoxon Signed-Rank menunjukkan adanya perbedaan performa yang signifikan antara algoritma Jaro-Winkler dan Levenshtein. Dengan demikian, perbedaan kinerja yang diperoleh bukan hanya bersifat numerik, tetapi juga signifikan secara statistik.

5.2 Saran

1. Penelitian selanjutnya dapat menambahkan pendekatan hybrid, misalnya menggabungkan string similarity dengan metode semantic similarity seperti TF-IDF, cosine similarity, atau embedding berbasis model bahasa, untuk meningkatkan ketepatan deteksi kemiripan judul.
2. Sistem yang dikembangkan selanjutnya sebaiknya diintegrasikan dengan antarmuka pengguna dan basis data yang terstruktur sehingga dapat digunakan langsung oleh admin akademik atau dosen pembimbing dalam proses verifikasi judul.



DAFTAR PUSTAKA

- Assaf, R., & Saad, S. (2022). A comparative analysis of string similarity algorithms for record linkage. *Journal of Information and Data Management*, 13(1), 22–35.
- Azizah, N., Wibowo, A., & Hidayat, R. (2021). Evaluation of similarity threshold for record linkage using string matching techniques. *International Journal of Intelligent Systems and Applications*, 13(4), 45–52.
- Cohen, W. W., Ravikumar, P., & Fienberg, S. E. (2003). A comparison of string distance metrics for name-matching tasks. In *Proceedings of the 2003 International Conference on Information Integration on the Web (IIWeb-03)* (pp. 73–78). AAAI Press.
- Christen, P. (2012). *Data matching: Concepts and techniques for record linkage, entity resolution, and duplicate detection*. Springer. <https://doi.org/10.1007/978-3-642-31164-2>
- Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). SAGE Publications.
- GeeksforGeeks. (2024). *Jaro and Jaro-Winkler Similarity*. Diakses dari <https://www.geeksforgeeks.org/dsa/jaro-and-jaro-winkler-similarity/>
- Gomaa, W. H., & Fahmy, A. A. (2013). A survey of text similarity approaches. *International Journal of Computer Applications*, 68(13), 13–18. <https://doi.org/10.5120/11638-7118>
- Hidayat, R., Nurwanto, R., & Suyanto, S. (2020). Implementasi algoritma Jaro–Winkler dan Levenshtein untuk deteksi kemiripan judul skripsi. *Jurnal Teknologi dan Sistem Informasi*, 8(2), 75–82.
- Jaro, M. A. (1989). Advances in record-linkage methodology as applied to the 1985 census of Tampa, Florida. *Journal of the American Statistical Association*, 84(406), 414–420. <https://doi.org/10.2307/2289924>
- Jurafsky, D., & Martin, J. H. (2023). *Speech and language processing* (3rd ed.). Draft. Stanford University. <https://web.stanford.edu/~jurafsky/slp3/>
- Khadivi, E., Pourshahabi, V., & Aghaie, M. (2021). Improving entity matching using Jaro–Winkler similarity in large-scale datasets. *International Journal of Data Science and Analytics*, 12(4), 311–325.
- Khalid, M. (2022). Toward Efficient Similarity Search under Edit Distance on Large-Scale Data. *Information*, 13(10), 452.
- Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions and reversals. *Soviet Physics Doklady*, 10(8), 707–710.
- Maharani, D., Nugroho, A. S., & Wibowo, R. (2020). Optimasi nilai threshold pada algoritma string matching untuk record linkage. *Jurnal Teknologi*

- Informasi dan Ilmu Komputer(JTIK),7(3),497–505.
<https://doi.org/10.25126/jtiik.2020732403>
- Navarro, G. (2001). A guided tour to approximate string matching. *ACM Computing Surveys*,33(1),31–88.
<https://doi.org/10.1145/375360.375365>
- Pallant, J. (2020). *SPSS survival manual: A step by step guide to data analysis using IBM SPSS (7th ed.)*. McGraw-Hill Education.
- Powers, D. M. W. (2020). Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. *Journal of Machine Learning Technologies*, 2(1), 37–63.
- Rahman, F., Prasetyo, E., & Nurcahyo, A. (2022). Analisis kinerja algoritma string similarity untuk deteksi kemiripan judul skripsi menggunakan pendekatan hibrid. *Jurnal Teknologi dan Sistem Komputer*, 10(2), 89–96. <https://doi.org/10.14710/jtsiskom.10.2.2022.89-96>
- Sari, D. P., & Wibisono, B. (2021). Implementasi algoritma Jaro–Winkler dalam sistem deteksi kesamaan judul tugas akhir mahasiswa. *Jurnal Teknologi dan Sistem Informasi Bisnis*, 3(2), 45–52.
- Sari, N., Wijayanto, A., & Pratama, F. (2023). *GPU-based similarity metrics computation and machine learning approaches for string similarity evaluation in large datasets*. *Soft Computing Journal*.
- Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*,45(4),427–437.
<https://doi.org/10.1016/j.ipm.2009.03.002>
- Wang, J. (2020). Measurement of text similarity: A survey. *Information*, 11(9), 421.
- Wijaya, H., & Putra, M. (2022). *Jaro Winkler Algorithm for Measuring Similarity OnlineNews* *Jurnal Teknologi Informasi dan Ilmu Komputer*