

BAB 5

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil penerapan metode SVM untuk menganalisis sentimen data sintesis dan data asli, didapat beberapa kesimpulan yaitu:

1. Ekstraksi fitur dengan TF-IDF mengevaluasi pentingnya sebuah kata dalam suatu dokumen sedangkan *count vectorizer* melihat frekuensi sebuah kata dalam suatu dokumen.
2. Pada data sintesis terdapat 100 observasi dan dengan ekstraksi didapatkan sebanyak 161 kata yang memiliki nilai vektor. Lalu, pada data Twitter terdapat 500 observasi, dengan ekstraksi fitur didapatkan sebanyak 1991 kata yang memiliki nilai vektor. Performa yang dihasilkan model dengan menggunakan metode TF-IDF memberikan hasil yang lebih baik dari pada metode *count vectorizer* pada kedua set data.
3. Berdasarkan hasil pengujian klasifikasi SVM pada data sintesis setelah dilakukan modifikasi data, algoritma SVM dapat melakukan klasifikasi dengan baik. Kernel linear dengan metode ekstraksi fitur TF-IDF menghasilkan performa *F1-score* terbaik sebesar 80%. Algoritma SVM juga berkeja dengan baik pada data Twitter setelah dilakukan modifikasi data, di mana kernel RBF dengan metode ekstraksi fitur TF-IDF dan *count vectorizer* menghasilkan performa *F1-score* terbaik sebesar 73,47%. Hal ini dapat terjadi karena model klasifikasi SVM sudah mempelajari kata atau observasi yang terdapat pada data latih sehingga saat dilakukan pengujian dengan data uji dapat memberikan performa kinerja yang lebih baik dengan melakukan klasifikasi yang benar.

5.2 Saran

Berdasarkan analisis yang telah dilakukan untuk menganalisis sentimen, berikut ini merupakan saran untuk penelitian lebih lanjut, yaitu:

1. Penambahan daftar kata-kata positif dan negatif yang lebih spesifik untuk membedakan kedua sentimen.
2. Penggunaan kata dengan frekuensi yang lebih banyak dengan menambahkan jumlah observasi agar kata yang terdapat pada data uji juga terdapat pada data latih.

DAFTAR REFERENSI

- [1] Slamet, R., Gata, W., Novtariany, A., Hilyati, K., dan Jariyah, F. A. (2022) Analisis sentimen twitter terhadap penggunaan artis Korea Selatan sebagai brand ambassador produk kecantikan lokal. *INTECOMS: Journal of Information Technology and Computer Science*, **5**, 145–153.
- [2] Sharma, D. dan Sabharwal, M. (2019) Sentiment analysis for social media using svm classifier of machine learning. *Int J Innov Technol Exploring Eng (IJITEE)*, **8**, 39–47.
- [3] Nugroho, A., Witarto, A., dan Handoko, D. (2003) Application of support vector machine in bioinformatics. *Proceeding of Indonesian Scientific Meeting in Central Japan*.
- [4] Isnain, A. R., Supriyanto, J., dan Kharisma, M. P. (2021) Implementation of k-nearest neighbor (k-nn) algorithm for public sentiment analysis of online learning. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, **15**, 121–130.
- [5] Nguyen, H., Veluchamy, A., Diop, M., dan Iqbal, R. (2018) Comparative study of sentiment analysis with product reviews using machine learning and lexicon-based approaches. *SMU Data Science Review*, **1**, 7.
- [6] Abdillah, W. F., Premana, A., dan Bhakti, R. H. (2021) Analisis sentimen penanganan covid-19 dengan support vector machine: Evaluasi leksikon dan metode ekstraksi fitur. *Jurnal Ilmiah Intech: Information Technology Journal of UMUS*, **3**, 160–170.
- [7] Raza, G. M., Butt, Z. S., Latif, S., dan Wahid, A. (2021) Sentiment analysis on covid tweets: an experimental analysis on the impact of count vectorizer and tf-idf on sentiment predictions using deep learning models. *2021 International Conference on Digital Futures and Transformative Technologies (ICoDT2)*, pp. 1–6. IEEE.
- [8] Awad, M. dan Khanna, R. (2015) *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers*. Springer nature.
- [9] Al Qadi, L., El Rifai, H., Obaid, S., dan Elnagar, A. (2019) Arabic text classification of news articles using classical supervised classifiers. *2019 2nd International conference on new trends in computing sciences (ICTCS)*, pp. 1–6. IEEE.
- [10] 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**, 25–29.
- [11] Widaningrum, I., Mustikasari, D., Arifin, R., Tsaqila, S. L., dan Fatmawati, D. (2022) Algoritma term frequency–inverse document frequency (tf-idf) dan k-means clustering untuk menentukan kategori dokumen. *Prosiding SISFOTEK*, **6**, 145–149.
- [12] Khomsah, S. dan Aribowo, A. S. (2020) Text-preprocessing model youtube comments in Indonesian. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, **4**, 648–654.
- [13] Singh, P. (2018) *Machine Learning with PySpark: with Natural Language Processing and Recommender Systems*. Apress.

- [14] Kowalczyk, A. (2017) *Support Vector Machines Succinctly*. SynCFusion Inc.
- [15] Schwenke, C. dan Schering, A. (2014) *True Positives, True Negatives, False Positives, False Negatives*. Wiley Online Library.
- [16] Bird, S., Klein, E., dan Loper, E. (2009) *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*. O'Reilly Media, Inc.
- [17] Nurkholis, A., Alita, D., dan Munandar, A. (2022) Comparison of kernel support vector machine multi-class in ppkm sentiment analysis on twitter. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, **6**, 227–233.
- [18] Rosid, M. A., Fitriani, A. S., Astutik, I. R. I., Mulloh, N. I., dan Gozali, H. A. (2020) Improving text preprocessing for student complaint document classification using sastrawi. *IOP Conference Series: Materials Science and Engineering* 012017. IOP Publishing.
- [19] Berrar, D. (2018) Cross-Validation. *Encyclopedia of Bioinformatics and Computational Biology*, **1**, 542–545.

