

BAB 5

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil analisis penerapan algoritma SVM untuk mengklasifikasikan data ke dalam 2 kelas sentimen menggunakan data terstruktur dan tidak terstruktur, didapat beberapa kesimpulan yaitu:

1. Berdasarkan dari hasil 10 - *fold cross validation* algoritma klasifikasi SVM dapat melakukan klasifikasi teks dengan cukup baik, yaitu dengan mendapatkan nilai *f1-score* di atas 50% jika menggunakan kernel RBF dan kernel Linear.
2. Semakin besar jumlah data teks yang digunakan maka performa SVM dalam mengklasifikasikan teks akan mendapatkan hasil *f1-score* yang lebih tinggi. Baik itu menggunakan data terstruktur maupun tidak terstruktur.
3. Hasil *f1-score* yang didapat dari 3000 data teks terstruktur adalah 81.43% dan untuk data tidak terstruktur adalah 61.7%. Data terstruktur dan tidak terstruktur dapat diklasifikasikan dengan cukup baik menggunakan SVM. Perbedaan performa diakibatkan oleh data teks terstruktur memiliki lebih banyak kata-kata kunci pada setiap sentimennya sehingga memiliki hasil performa yang lebih baik dari data tidak terstruktur.

5.2 Saran

Berdasarkan hasil analisis penerapan algoritma SVM untuk mengklasifikasikan data teks, berikut ini merupakan saran untuk penelitian lebih lanjut, yaitu:

1. Pembuatan model klasifikasi SVM menggunakan data teks dengan lebih dari 2 sentimen, seperti sentimen positif, netral, dan negatif.
2. Analisis performa SVM pada data teks tingkatan dokumen atau paragraf.

DAFTAR REFERENSI

- [1] Graovac, J. dan Pavlovic-Lazetic, G. (2014) Language-independent sentiment polarity detection in movie reviews: a case study of english and spanish. *6th International Conference ICT Innovations*, pp. 13–22.
- [2] Kumar, S., Morstatter, F., dan Liu, H. (2014) *Twitter data analytics*. Springer.
- [3] Li, X. dan Liu, C. (2020) Comparison of machine learning models for sentimental analysis of hotel reviews. *IOP Conference Series: Materials Science and Engineering* 012029. IOP Publishing.
- [4] Kowsari, K., Jafari Meimandi, K., Heidarysafa, M., Mendu, S., Barnes, L., dan Brown, D. (2019) Text classification algorithms: A survey. *Information*, **10**, 150.
- [5] Hussein, D. M. E.-D. M. (2018) A survey on sentiment analysis challenges. *Journal of King Saud University-Engineering Sciences*, **30**, 330–338.
- [6] James, G., Witten, D., Hastie, T., dan Tibshirani, R. (2013) *An introduction to statistical learning*. Springer.
- [7] Kowalczyk, A. (2017) *Support Vector Machines Succinctly*. SynCFusion, Marrisville.
- [8] Chicco, D. dan Jurman, G. (2020) *The advantages of the Matthews Correlation Coefficient (MCC) over F1 score and accuracy in binary classification evaluation*. Springer.
- [9] Tharwat, A. (2020) *Classification assessment methods*. Emerald Publishing Limited.
- [10] Elbagir, S. dan Yang, J. (2019) Twitter sentiment analysis using natural language toolkit and vader sentiment. *Proceedings of the international multiconference of engineers and computer scientists* 16.
- [11] Koto, F. dan Rahmaningtyas, G. Y. (2017) Inset lexicon: Evaluation of a word list for indonesian sentiment analysis in microblogs. *2017 International Conference on Asian Language Processing (IALP)*, pp. 391–394. IEEE.
- [12] Cruz, C. A. A. dan Balahadia, F. F. (2021) Analyzing public concern responses for formulating ordinances and laws using sentiment analysis through vader application. *International Journal of Computing Sciences Research*, **6**, 842–856.