

BAB 6

KESIMPULAN DAN SARAN

6.1 Kesimpulan

Dari penelitian yang sudah dilakukan, ada beberapa kesimpulan di antaranya:

1. Pembuatan fitur dari teks berita dapat menggunakan *Vector Space Model* untuk merepresentasikan dokumen menjadi sebuah vektor yang melihat tingkat kedekatan atau kesamaannya dengan cara pembobotan *term*.
2. Reduksi Dimensi pada *Vector Space Model* (VSM) dilakukan dengan menerapkan *Singular Value Decomposition* (SVD) yang merupakan sebuah teknik dalam metode *Latent Semantic Indexing* (LSI). Reduksi dimensi dilakukan dengan menentukan jumlah dimensi matriks yang tersisa yang didapatkan dari jumlah sebuah nilai *singular value*, yang akan terbentuk matriks baru yang didapatkan dari perkalian matriks *document-term*. Reduksi dimensi matriks mendapatkan model yang efektif yang dapat merepresentasikan hubungan konsep(kategori) antar kata dan dokumen.
3. Secara performa keseluruhan, algoritma *Naive Bayes* dengan fitur TF-IDF, algoritma *Support Vector Machine* dengan fitur TF-IDF dan algoritma *Support Vector Machine* dari fitur LSI bekerja sangat baik dibanding algoritma lain. Karena dari kedua fitur TFIDF maupun LSI nilai dari *F-Measure* dan *Accuracy* mencapai 90% lebih, dengan nilai *F-Measure* pada TFIDF-Naive Bayes 92.696%, TFIDF-SVM 92.276%, dan LSI-SVM 92.982%. Untuk nilai *Accuracy* pada TFIDF-Naive Bayes 96.97%, TFIDF-SVM 96.922% dan LSI-SVM 95.918%. Penelitian ini membuktikan bahwa metode TFIDF-Naive Bayes dan SVM dari tingkat klasifikasi ketepatan, keberhasilan dan akurasi data dalam mengelompokkan suatu dataset berita yang dipakai cukup optimal, karena terdapat banyak kategori artikel yang terklasifikasi dengan benar oleh algoritma klasifikasinya sehingga mencap nilai yang tinggi dan akurat.
4. Metode *Latent Semantic Indexing* (LSI) bekerja baik ketika digunakan pada algoritma *K-Nearest Neighbor* (KNN) karena didapatkan penaikan nilai *F-Measurement* maupun *Accuracy*. Namun, pada hasil TFIDF-KNN tidak terlalu bagus meskipun sudah mencari nilai *k* yang paling terbaik, hasil akurasi yang dihasilkan relatif kecil dibandingkan dengan metode lainnya yaitu sekitar 83% dikarenakan KNN hanya memperhatikan tetangga disekitar *k-data training* yang paling dekat jaraknya dengan data *testing*, maka banyak kategori artikel yang tidak terklasifikasi baik pada artikel tersebut, sehingga kurang efektif dalam mengklasifikasikan berita yang membutuhkan pengaruh fitur dalam penentuan kelas.

6.2 Saran

Saran untuk penelitian selanjutnya adalah seperti berikut.

1. Program berhasil dibuat dan bekerja dengan cukup baik. Namun pada pengujian performa mengalami kendala pada proses *text preprocessing* yang sangat lama apabila menggunakan dataset dengan jumlah yang banyak atau besar. Ini disebabkan karena pembobotan *term* dalam banyak dokumen untuk membentuk matriks *document-term*.
2. Dataset atau dokumen berita yang digunakan masih cukup sedikit dengan berjumlah 750

dokumen dalam bentuk .csv. Hal tersebut dikarenakan dataset yang terlalu besar takut membuat perangkat lunak atau *tool* yang digunakan tidak sanggup untuk merespon.

3. Metode dalam pengujian diperbanyak lagi agar perbandingannya lebih bagus dan metode mana yang lebih tepat untuk klasifikasi dokumen berita.
4. Dalam penelitian selanjutnya dapat menggunakan data lainnya, karena penggunaan dataset artikel berita belum terlalu optimal dalam pengklasifikasian.

DAFTAR REFERENSI

- [1] Hasan, M. M. dan Matsumoto, Y. (2013) Document clustering: Before and after the singular value decomposition. Technical Report UU-CS-2013-007. Nara Institute of Science and Technology, Nara, Japan.
- [2] Kou, G. dan Peng, Y. (2015) An application of latent semantic analysis for text categorization. *International Journal of Computers Communications & Control*, Chengdu, China, June, pp. 357–369. Springer.
- [3] Josi, A., Abdillah, L. A., dan Suryayusra (2014) Penerapan teknik web scraping pada mesin pencari artikel ilmiah. Technical report. Universitas Bina Darma, Palembang.
- [4] Viranda, L., Hidayat, D. R., dan Yudhapramesti, P. (2018) Manajemen produk media Kompas.com untuk bersaing di era generasi milenial. *Kajian Jurnalisme*, **2**, 61–78.
- [5] Murdiowinata, A. dan Supratman, E. (2020) Backpropagation untuk menentukan peramalan hasil produksi buah sawit pada PT Gemilang Cahaya Mentari. *Bina Darma Conference on Computer Science*, **2**, 172–181.
- [6] Rachmawati, T. S. dan Agustine, M. (2021) Keterampilan literasi informasi sebagai upaya pencegahan hoaks mengenai informasi kesehatan di media sosial. *Jurnal Kajian Informasi & Perpustakaan*, **9**, 99–114.
- [7] Asistiyasari, A., Sudarsono, B., dan Faddilah, U. (2021) Klasifikasi berita hoaks menggunakan algoritma vector space model classification of hoax news using the vector space model algorithm. *Jurnal Informatika & Teknologi*, **2**, 114–121.
- [8] Ardiansyah, A. H., Kartika, K. P., dan Budiman, S. N. (2021) Penerapan latent semantic indexing pada sistem temu balik informasi pada undang-undang pemilu berdasarkan kasus. *Jurnal MNEMONIC*, **4**, 64–70.
- [9] Novalia, V., Goejantoro, R., dan Sifriyani (2020) Perbandingan metode klasifikasi naive bayes dan k-nearest neighbor (studi kasus : Status kerja penduduk di kabupaten kutai kartanegara tahun 2018). *Jurnal EKSPONENSIAL*, **11**, 159–166.
- [10] Hermanto, Mustopa, A., dan Kuntoro, A. Y. (2020) Algoritma klasifikasi naive bayes dan support vector machine dalam layanan komplain mahasiswa. *Jurnal ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER*, **5**, 211–220.
- [11] Deviyanto, A. dan Wahyudi, M. D. R. (2018) Penerapan analisis sentimen pada pengguna twitter menggunakan metode k-nearest neighbor. *Jurnal Informatika Sunan Kalijaga*, **3**, 1–13.
- [12] Kusriani dan Luthfi, E. T. (2009) *Algoritma Data Mining*, 1st edition. Andi Offset, Yogyakarta.
- [13] Septian, G., Susanto, A., dan Shidik, G. F. (2017) Indonesian news classification based on nabana. *Application for Technology of Information and Communication*, **2**, 175–180.
- [14] Asyarie, A. D. dan Pribadi, A. W. (2019) Automatic news articles classification in Indonesian language by using naive bayes classifier method. *Journal of ERPAS*, **3**, 658–662.

-
- [15] Lidya, S. K. (2014) Sentiment analysis pada teks bahasa indonesia menggunakan support vector machine (svm) dan k-nearest neighbor (knn), . Sumatera Utara, August, pp. 1–8. Universitas Sumatera Utara.
- [16] D, D. M., C, S., dan Ganesh, A. (2016) Sentiment analysis:a comparative study on different approaches. *Procedia Computer Science*, **10**, 44–49.
- [17] Ginantra, N. L. W. S. R., Arifah, F. N., dan Wijaya, A. H. (2021) *Data Mining dan Penerapan Algoritma*, 1st edition. Yayasan Kita Menulis, Medan.
- [18] Han, J., Kamber, M., dan Pei, J. (2012) *Data Mining: Concepts and Techniques*, 3rd edition. Morgan Kaufmann, Waltham.
- [19] Octaviani, P. A., Wilandari, Y., dan Ispriyanti, D. (2014) Penerapan metode klasifikasi support vector machine (svm) pada data akreditasi sekolah dasar (sd) di kabupaten magelang. *Jurnal Gaussian*, **3**, 811–820.
- [20] Claudy, Y. I., Perdana, R. S., dan Fauzi, M. A. (2018) Klasifikasi dokumen twitter untuk mengetahui karakter calon karyawan menggunakan algoritme k-nearest neighbor (knn). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, **2**, 2761–2765.
- [21] Siahaan, S. P., Dr. Eng. Willy Anugrah Cahyadi, M., S.T., dan Dr. Eng. Willy Anugrah Cahyadi, M., S.T. (2020) Pemantauan dan pengontrolan sistem pemupukan budidaya tanaman tomat menggunakan convolutional neural network monitoring and controlling fertilization system for tomato cultivation using convolutional neural network. *Journal of Engineering*, **7**, 1–15.
- [22] Zhai, C. dan Massung, S. (2016) *Text Data Management and Analysis: A Practical Introduction to Information Retrieval and Text Mining*, 12th edition. ACM Books, New York.
- [23] Kumar, V. dan Minz, S. (2014) Feature selection: A literature review. *Journal of Smart Computing Review*, **4**, 211–229.
- [24] Wall, M. E., Rechtsteiner, A., dan Rocha, L. M. (2003) Singular value decomposition and principal component analysis. *Journal of Practical Approach to Microarray Data Analysis*, **5**, 91–109.
- [25] Lu, W. Z., Wang, W. J., Wang, X. K., Yang, S. H., dan Lam, J. C. (2004) Potential assessment of a neural network model with pca/rbf approach for forecasting pollutant trends in mong kok urban air, hong kong. *Journal of Environmental*, **96**, 79–87.
- [26] Sedghpour, A. S. dan Sedghpour, M. R. S. (2020) Web document categorization using naive bayes classifier and latent semantic analysis. *Latent Semantic Analysis*, **1**, 3.
- [27] Steinberger, J. dan Jezek, K. (2004) Using latent semantic analysis in text summarization and summary evaluation, . Ceko, January, pp. 93–100. InProc ISIM.
- [28] Zhao, B. (2017) Web scraping. Bagian dari Laurie A. Schintler, C. L. M. (ed.), *Encyclopedia of Big Data*. Springer International Publishing.
- [29] Pedregosa, F., Varoquaux, G., dan Gramfort, A. (2011) Scikitlearn: Machine learning in python. *Journal of Machine Learning Research*, **85**, 2825–2830.
- [30] Dwimarcayani, D., Badriyah, T., dan Karlita, T. (2019) Classification on category of public responses on television program using naive bayes method. *Journal of International Electronics Symposium (IES)*, **19**, 225–231.
- [31] Purnamasari, K. dan Suwardi, I. (2018) Rule-based part of speech tagger for indonesian language, . Ceko, August, pp. 1–4. IOP Publishing.

-
- [32] 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. *Journal of IOP Materials Science and Engineering*, **1**, 1–6.