

BAB 6

KESIMPULAN DAN SARAN

6.1 Kesimpulan

Berdasarkan hasil penelitian yang sudah dilakukan pada Tugas Akhir ini, berikut ini adalah beberapa kesimpulan yang dapat diambil:

1. Data teks telah berhasil dimodelkan ke dalam bentuk vektor dengan elemen-elemen yang mewakili derajat keanggotaan antar kluster. Dengan demikian, dimensi fitur menjadi lebih kecil dan lebih padat (*dense*). Namun, perlu dilakukan normalisasi pada nilai-nilai tersebut agar nilai antar fitur pada data dengan label yang berbeda juga berbeda. Oleh karena itu, fitur untuk model klasifikasi adalah derajat keanggotaan antar kluster yang telah dinormalisasi menggunakan *min-max scaling*.
2. Model klasifikasi Multinomial Naive Bayes dimana data pelatihan dan pengujiannya adalah *dataset* TF-IDF memiliki performa yang lebih unggul dibandingkan dengan model klasifikasi yang data pelatihan dan pengujiannya adalah *dataset* yang di reduksi dimensi menggunakan algoritma Fuzzy C-Means sehingga fitur dari data menjadi derajat keanggotaan antar kluster dan dilanjutkan dengan proses normalisasi menggunakan *min-max scaling*. Nilai akurasi adalah 90% : 62%.

Tetapi untuk performa model Logistic Regression, dimana data pelatihan dan pengujiannya adalah *dataset* TF-IDF memiliki performa yang sedikit lebih tinggi dibandingkan dengan yang data pelatihan dan pengujiannya adalah *dataset* yang di reduksi dimensi menggunakan algoritma Fuzzy C-Means sehingga fitur dari data menjadi derajat keanggotaan antar kluster dan dilanjutkan dengan proses normalisasi menggunakan *min-max scaling*. Nilai akurasi adalah 88% : 87%.

3. Perangkat lunak berbasis *website* yang dibangun untuk melakukan prediksi terhadap data baru yang diunggah oleh pengguna telah dapat digunakan. Perangkat lunak telah diuji fungsionalitasnya dan berfungsi dengan baik. Model *machine learning* Logistic Regression yang telah dilatih dan diuji menggunakan *dataset* hasil direduksi dimensi algoritma Fuzzy C-Means lalu dilanjutkan dengan proses normalisasi *min-max scaling* pada setiap fiturnya, telah diimplementasikan pada perangkat lunak.

6.2 Saran

Berikut ini adalah beberapa saran yang dapat dilakukan untuk pengembangan selanjutnya:

1. Algoritma *fuzzy clustering* yang diimplementasikan untuk mereduksi dimensi data teks pada penelitian ini adalah algoritma dengan pendekatan jarak yaitu *Fuzzy C-Means* (FCM). Untuk pengembangan lebih lanjut, dapat menggunakan algoritma dengan pendekatan lain seperti pendekatan probabilistik.
2. Data teks yang digunakan pada eksperimen tugas akhir merupakan data ulasan. Untuk pengembangan lebih lanjut, dapat menggunakan data dengan topik lain seperti data teks artikel berita.
3. Pemodelan awal data hanya TF-IDF. Untuk pengembangan lebih lanjut, dapat memanfaatkan *word embedding* pada vektor TF-IDF lalu dilakukan reduksi dimensi dengan algoritma klusterisasi.

DAFTAR REFERENSI

- [1] Pustejovsky, J. dan Stubbs, A. (2012) *Natural Language Annotation for Machine Learning*, 1st edition. O'Reilly, Beijing.
- [2] Jurafsky, D. dan Martin, J. H. (2008) *Speech and Language Processing*, 3rd edition. Pearson Prentice Hall, USA.
- [3] Patel, D. dan Chhinkaniwala, H. (2018) Fuzzy logic-based single document summarisation with improved sentence scoring technique. *International Journal of Knowledge Engineering and Data Mining*, **5**, 125–138.
- [4] Peng, H., Pavlidis, N., Eckley, I., dan Tsalamani, I. (2018) Subspace clustering of very sparse high-dimensional data. *2018 IEEE International Conference on Big Data (Big Data)*, **28**, 3780–3783.
- [5] Karami, A. (2017) Taming wild high dimensional text data with a fuzzy lash. *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*, **15**, 518–522.
- [6] Aggarwal, C. C. dan Zhai, C. (2012) An introduction to text mining. *Mining Text Data*, **8**, 1–10.
- [7] Crain, S. P., Zhou, K., Yang, S.-H., dan Zha, H. (2012) Dimensionality reduction and topic modeling: From latent semantic indexing to latent dirichlet allocation and beyond. *Mining Text Data*, **11**, 129–161.
- [8] Feldman, R. dan Sanger, J. (2007) *The Text Mining Handbook*, 1st edition. Ronan Feldman and James Sanger, New York.
- [9] Han, J., K., M., dan Pei, J. (2012) *Data Mining Concepts and Techniques*, 3rd edition. Morgan Kaufmann, 225 Wyman Street, Waltham, MA 02451, USA.
- [10] Mueller, J. P. dan Massaron, L. (2021) *Machine Learning for Dummies*, 2nd edition. John Wiley and Sons, Inc., Hoboken, New Jersey.
- [11] James, G., Witten, D., Hastie, T., dan Tibshirani, R. (2013) *An Introduction to Statistical Learning with Applications in R*, 1st edition. Springer, New York City, USA.
- [12] Bede, B. (2013) *Mathematics of Fuzzy Sets and Fuzzy Logic*, 1st edition. Springer-Verlag, Berlin.
- [13] Prasetyo, E. (2014) *Data Mining Mengolah Data Menjadi Informasi Menggunakan Matlab*, 1st edition. Andi Offset, Yogyakarta.
- [14] Singh, A., Y., A., dan Rana, A. (2013) K-means with three different distance metrics. *Computer Applications*, **67**, 975–8887.
- [15] Davies, D. dan Bouldin, D. (1979) A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **2**, 224–227.

-
- [16] Huang, M., Zhu, X., Hao, Y., Payan, D. G., Qu, K., dan Li, M. (2004) Discovering patterns to extract protein–protein interactions from full texts. *Bioinformatics*, **20**, 3604–3612.
- [17] Brodlie, K., Osorio, R. S. A., dan Lopes, A. (2012) A review of uncertainty in data visualization. *White Rose Research Online*, **12**, 81–110.
- [18] Moertini, V. S. dan Adithia, M. T. (2020) *Pengantar Data Science dan Aplikasinya bagi Pemula*, 1st edition. UNPAR Press, Bandung.
- [19] Kirk, A. (2012) *Data Visualization: A Successful Design Process*, 1st edition. Packt Publishing, United Kingdom.
- [20] Cao, N. dan Cui, W. (2016) *Introduction to Text Visualization*, 1st edition. Atlantis Press, Paris.
- [21] Bird, S., Klein, E., dan Loper, E. (2009) *Natural Language Processing with Python*, 1st edition. O’Reilly Media, Inc., California.
- [22] Maas, A. L., Daly, R. E., Pham, P. T., Huang, D., Ng, A. Y., dan Potts, C. (2011) Learning word vectors for sentiment analysis. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*.
- [23] Maas, A. L., Daly, R. E., Pham, P. T., Huang, D., Ng, A. Y., dan Potts, C. (2011) Learning word vectors for sentiment analysis. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, Portland, Oregon, USA, June, pp. 142–150. Association for Computational Linguistics.
- [24] Karami, A. (2019) Application of fuzzy clustering for text data dimensionality reduction. *Int. J. Knowledge Engineering and Data Mining*, **6**, 1–19.