

BAB 6

KESIMPULAN DAN SARAN

Pada bab ini akan dibahas mengenai kesimpulan yang didapatkan dalam penelitian ini. Selain membahas mengenai kesimpulan dari penelitian ini, terdapat bagian saran yang dapat digunakan sebagai referensi penelitian lebih lanjut.

6.1 Kesimpulan

Berdasarkan penelitian yang sudah dilakukan di mana penelitian ini terdiri dari proses pengumpulan data, penyiapan data, perancangan model, dan pembangunan perangkat lunak, didapatkan kesimpulan sebagai berikut:

1. Telah dilakukan pengumpulan data *Twitter* berbahasa Indonesia tentang sekolah tatap muka di Indonesia. Teknik yang digunakan untuk dapat melakukan penarikan data ini adalah dengan melakukan implementasi API *Twitter*, *tweepy*, dengan melakukan konfigurasi bahasa yang digunakan adalah Bahasa Indonesia dengan kueri "sekolah tatap muka". Untuk pengumpulan data *Twitter*, didapatkan *tweet* sebanyak 17913 buah *tweet*.
2. Telah dilakukan penyiapan data pesan *Twitter* untuk analisis lebih lanjut dengan menggunakan berbagai teknik dalam NLP dan teknik pre-processing data. Adapun teknik yang digunakan adalah *data cleaning*, filterisasi, tokenisasi, dan normalisasi dengan algoritma Nazief-Adriani. Hasil dari penyiapan data adalah data yang sudah bersih yang siap untuk dianalisis lanjut.
3. Telah dipelajari mengenai konsep dasar klasifikasi *decision tree* sebagai klasifikator untuk proses perancangan model. Model akan kemudian digunakan untuk melakukan prediksi sentimen. Selain itu, proses ini dilakukan analisis data dengan menggunakan berbagai jenis visualisasi data yang tepat.
4. Telah dipelajari mengenai pembangunan perangkat lunak dengan menggunakan *library* yang sudah disediakan oleh *Python*. Untuk *library* yang digunakan untuk pembangunan perangkat lunak untuk melakukan analisis sentimen adalah menggunakan *pandas*, *matplotlib.pyplot*, *sklearn*, *PySastrawi*, *nlTK*, *pickle*, dan *wordcloud*. Selain perangkat lunak untuk analisis sentimen, terdapat bagian pembangunan perangkat lunak berbasis GUI dengan memanfaatkan *tools* *PAGE* dan *tkinker*.

6.2 Saran

Penelitian ini memang masih jauh dari kata sempurna. Masih terdapat kesalahan maupun kekeliruan yang terjadi dalam penelitian ini. Oleh karena itu, penelitian ini masih dapat dilanjutkan dengan berbagai saran yang penulis berikan. Berikut merupakan saran penulis yang dapat dijadikan preferensi untuk penelitian selanjutnya:

- Mencoba mengimplementasikan *typo correction* atau perbaikan kata untuk kata-kata yang salah ketik. Hal ini didasari oleh pengguna *Twitter* yang sering melakukan kesalahan pengetikan. Akan tetapi perlu diperhatikan, *typo correction* berpotensi untuk mengubah struktur kalimat. Misalnya pada kasus kata 'yg', *typo correction* yang paling memungkinkan keluar adalah kata 'ya' dibandingkan kata 'yang'.

- Melakukan implementasi *polarity score* yang lebih beragam. Dalam penelitian ini, nilai polaritas yang digunakan hanya sebatas nilai 1 untuk positif dan -1 untuk nilai negatif. Permasalahan teknik ini adalah kata yang bersentimen belum tentu memiliki polaritas sentimen yang sama. Kata ‘bagus’, ‘oke’, ‘sempurna’, dan ‘baik’ merupakan kata sentimen positif. Akan tetapi, polaritas sentimen keempat kata tersebut berbeda. Kata ‘oke’ memiliki polaritas positif yang paling rendah karena kata ‘oke’ menunjukkan bahwa terdapat aksi yang sudah dilakukan namun berada di level minimum. Kata ‘sempurna’ memiliki polaritas positif paling tinggi karena kata tersebut memiliki arti tidak ada cacat atau kekurangan sama sekali.
- Mencoba parameter *test-split* selain 70% sebagai *train* dan 30% untuk *test*. Dalam penelitian ini, hanya digunakan *test-split* dengan nilai 70-30. Jika menggunakan nilai parameter 80-20 ataupun 75-25 akan memiliki nilai akurasi, *precision*, *recall*, dan *f-score* yang berbeda.
- Mencoba parameter *decision tree* lainnya selain yang sudah dicoba dalam penelitian ini. Dalam penelitian ini, *decision tree* hanya menggunakan parameter **criterion** dengan jenis *entropy* dan **max_depth** sebesar 135. Terdapat parameter lain yang diindikasikan dapat mengubah kualitas model yang dibangun. Salah satu *parameter* dalam *decision tree* yang dapat dicoba adalah **minNumObjects** atau jumlah minimum objek yang ditampung oleh suatu *leaf*. Dengan pengaturan minimum jumlah objek yang ditampung oleh *leaf*, *tree* yang dihasilkan akan lebih optimal.
- *Dataset* dapat ditambahkan lagi sebab dalam proses prediksi analisis sentimen, beberapa kata yang mengandung sentimen tidak terdeteksi pada kategori sentimen positif atau negatif. Hal ini disebabkan oleh kurangnya data yang berdampak pada pembuatan model *tree* yang dihasilkan. Pada kasus penelitian ini, kata “benci” dikategorikan sebagai netral karena frekuensi kata “benci” pada dataset memiliki jumlah yang sedikit.

DAFTAR REFERENSI

- [1] Fitri, V. A., Andreswari, R., dan Hasibuan, M. A. (2019) Sentiment analysis of social media twitter with case of anti-lgbt campaign in indonesia using naïve bayes, decision tree, and random forest algorithm. *Procedia Computer Science*, **161**, 765–772.
- [2] Géron, A. (2019) *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems*. " O'Reilly Media, Inc."
- [3] Han, J., Pei, J., dan Kamber, M. (2011) *Data mining: concepts and techniques*. Elsevier, Wyman Street, Waltham, MA 02451, USA.
- [4] Ivan (2018) Penambangan pesan twitter untuk memahami sentimen masyarakat terhadap perubahan harga pangan di indonesia. Skripsi. Universitas Katolik Parahyangan, Bandung, Indonesia.
- [5] Jurafsky, D. (2020) *Speech & Language Processing 3rd Edition*. Pearson Education India, Colorado.
- [6] Hasrul, T. (2020) Analisis kesuksesan film dengan data mining. Skripsi. Universitas Katolik Parahyangan, Bandung, Indonesia.
- [7] Zong, C., Xia, R., dan Zhang, J. (2021) *Text Data Mining*. Springer, China.
- [8] Suciadi, J. (2001) Studi analisis metode-metode parsing dan interpretasi semantik pada natural language processing. *Jurnal Informatika*, **2**, 13–22.
- [9] Pramudita, H. R. (2014) Penerapan algoritma stemming nazief & adriani dan similarity pada penerimaan judul thesis. *Data Manajemen dan Teknologi Informasi (DASI)*, **15**, 15.
- [10] Asian, J., Williams, H. E., dan Tahaghoghi, S. M. (2005) Stemming indonesian. *Proceedings of the Twenty-eighth Australasian conference on Computer Science-Volume 38*, 319 Darlinghurst, NSW 2010, Australia, January, pp. 307–314. Citeseer Australian Computer Society, Inc.
- [11] Adriani, M., Asian, J., Nazief, B., Tahaghoghi, S. M., dan Williams, H. E. (2007) Stemming indonesian: A confix-stripping approach. *ACM Transactions on Asian Language Information Processing (TALIP)*, **6**, 1–33.
- [12] Mustakim (2015) *Bentuk dan Pilihan Kata*. Kementrian Pendidikan dan Kebudayaan, Jakarta.
- [13] Liu, X., Li, B., Shen, D., Cao, J., dan Mao, B. (2017) Analysis of grain storage loss based on decision tree algorithm. *Procedia computer science*, **122**, 130–137.
- [14] Hendrawan, A., Winarti, T., dan Indriyawati, H. (2020) Pengembangan stemming untuk artikel berbahasa indonesia. Technical Report 227. Universitas Semarang, Semarang, Indonesia.
- [15] Rachman, F. F. dan Pramana, S. (2020) Analisis sentimen pro dan kontra masyarakat indonesia tentang vaksin covid-19 pada media sosial twitter. *Indonesian of Health Information Management Journal (INOHIM)*, **8**, 100–109.

- [16] Lalithamani, N., Thati, L. S., dan Adhikesavan, R. (2014) Sentence level sentiment polarity calculation for customer reviews by considering complex sentential structures. *IJRET: International Journal of Research in Engineering and Technology*, **3**, 433–438.
- [17] Liu, B., Hu, M., dan Cheng, J. (2005) Opinion observer: analyzing and comparing opinions on the web. *Proceedings of the 14th international conference on World Wide Web*, pp. 342–351.
- [18] Wahid, D. H. dan Azhari, S. (2016) Peringkasan sentimen esktraktif di twitter menggunakan hybrid tf-idf dan cosine similarity. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, **10**, 207–218.