

## BAB V KESIMPULAN DAN SARAN

### 5.1 Kesimpulan

Penelitian ini bertujuan untuk melihat performa model klasifikasi Naïve Bayes dengan mengkombinasikan TF-IDF membangun aplikasi pendeteksi cyberbullying dan mengukur akurasi terhadap pemodelannya, menggunakan metode penelitian SDLC waterfall untuk penelitian ini.

Berdasarkan temuan penelitian, kesimpulan yang dapat diambil adalah menggunakan metode pengembangan SDLC *waterfall*, sistem telah berhasil dibangun secara bertahap dan terstruktur. Tahapan ini meliputi Analisis Kebutuhan dengan mengumpulkan komentar Instagram yang mengandung *Cyberbullying* sebagai datasetnya untuk dianalisa diikuti tahap Desain Sistem yang merancang arsitektur setiap tahapannya, mencakup *preprocessing*, TF-IDF, dan pemodelan *Naïve Bayes* lalu Implementasi Membangun sistem sesuai dengan desain yang telah dibuat, mencakup implementasi *preprocessing*, TF-IDF dan pemodelan *Naïve Bayes* diakhiri dengan Pengujian : melakukan pengujian *Cross Validation* terhadap pemodelan dan sistem berjalan sesuai dengan harapan.

Dari data yang telah diambil dengan jumlah kalimat sebanyak 650 dalam bentuk teks, dilakukan analisis menggunakan metode *Naïve Bayes* dan TF IDF, analisis melalui beberapa langkah yang sistematis dan terstruktur dimulai dari pengumpulan data lalu dilanjutkan dengan *preprocessing* untuk memastikan data yang dikumpulkan sesuai dengan kebutuhan agar dataset dapat digunakan pada langkah pemodelan. Pada langkah pengujian, temuan penelitian dari penelitian ini pada langkah pembobotan kata menggunakan TF-IDF memunculkan fitur sebanyak 2000, lalu dilakukan pengujian *Cross Valid*

*ation* untuk mencegah adanya *Overfitting* dan mengukur performa model. Setelah itu dilakukan pemodelan menggunakan *Naïve Bayes* dengan proporsi data latih dan data uji sebesar 60:40, 70:30, 80:20 yang masing masing memiliki akurasi 79%, 98%, dan 80%, namun pemodelan masih tetap *overfitting*.

Merancang dan membangun sistem identifikasi *cyberbullying* menggunakan metode *Waterfall* dengan algoritma *Naïve Bayes* dan TF-IDF melibatkan proses yang sistematis dari analisis kebutuhan hingga evaluasi akhir. Sistem yang dihasilkan diharapkan mampu menganalisis sentimen dengan akurat, memberikan wawasan berharga tentang opini dan persepsi yang dapat digunakan untuk berbagai tujuan, seperti peningkatan produk, atau strategi pemasaran.

## 5.2 Saran

Penelitian ini terdapat banyak kekurangan dalam pembuatan laporan, bahkan memiliki kekurangan dalam pembuatan sistem identifikasi *cyberbullying*. Sehingga perlu adanya saran untuk mengembangkan laporan dan sistemnya. Maka penelitian selanjutnya diharapkan :

1. Memungkinkan untuk Pengembangan dan penambahan fitur agar dapat menggambarkan temuan analisa yang telah didapat.
2. Menggunakan data yang lebih besar agar performa pemodelan lebih baik, agar dapat memperbaiki *overfitting* yang ada pada penelitian kali ini dengan cara memperbaiki proses ekstraksi fitur, dengan mencoba *n-grams* (*bigrams*, *trigrams*) selain *unigram*, atau menggunakan teknik *embedding* yang lebih canggih seperti *Word2Vec*, *GloVe*, atau *BERT*.
3. Melakukan *Hyperparameter* dan *Pipeline Optimization* pada semua langkah agar terintegrasi dengan baik dan terhindar dari *overfitting*.

## DAFTAR PUSTAKA

Asmara, R. A., Andjani, B. S., Rosiani, U. D., & Choirina, P. (2018). Klasifikasi Jenis Kelamin Pada Citra Wajah Menggunakan Metode Naïve Bayes. *Jurnal Informatika Polinema*, 4(3), 212-212.

Dataindonesia.id. (2022, 13 Juni). Tingkat Penetrasi Internet di Indonesia Menurut Usia. Diakses pada 12 januari 2023, dari <https://dataindonesia.id/digital/detail/remaja-paling-banyak-gunakan-internet-di-indonesia-pada-2022>

Diwandanu, M. T., & Wisudawati, L. M. (2023). ANALISIS SENTIMEN TERHADAP TWIT MAXIM PADA TWITTER MENGGUNAKAN R PROGRAMMING DAN K NEAREST NEIGHBORS. *Jurnal Ilmiah Informatika Komputer*, 28(1), 1-16.

Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv preprint arXiv:2203.05794*.

Hasanah, U., Astuti, T., Wahyudi, R., Rifai, Z., & Pambudi, R. A. (2018, November). An experimental study of text preprocessing techniques for automatic short answer grading in Indonesian. In *2018 3rd International Conference on Information Technology, Information System and Electrical Engineering (ICITISEE)* (pp. 230-234). IEEE.

Karyanti, M. P., & Aminudin, S. P. (2019). *Cyberbullying & Body Shaming*. Penerbit K-Media.

Khadafi, Madonna Al, Kartika, Kurnia Paranitha, & Febrinita, Filda (2022). Penerapan Metode *Naïve Bayes* Classifier dan *Lexicon Based* Untuk Analisis Sentimen *Cyberbullying* pada BPJS. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 6(2), 725-733, ISSN 2598-828X, LPPM ITN Malang, <https://doi.org/10.36040/jati.v6i2.5633>

Khaira, U., Johanda, R., Utomo, P. E. P., & Suratno, T. (2020). Sentiment analysis of cyberbullying on twitter using SentiStrength. *Indonesian Journal of Artificial Intelligence and Data Mining*, 3(1), 21-27.

Kosasih, R., & Alberto, A. (2021). Analisis Sentimen Produk Permainan Menggunakan Metode TF-IDF Dan Algoritma K-Nearest Neighbor. *InfoTekJar: Jurnal Nasional Informatika dan Teknologi Jaringan*, 6(1), 134-139.

Krstinić, D., Braović, M., Šerić, L., & Božić-Štulić, D. (2020). Multi-label classifier performance evaluation with confusion matrix. *Computer Science & Information Technology*, 1.

Langarizadeh, M., & Moghbeli, F. (2016). Applying naive bayesian networks to disease prediction: a systematic review. *Acta Informatica Medica*, 24(5), 364.

Lets End Bullying. (2013). (Priyatna.A.): Elex Media Komputindo.

Maulana, F. A., & Ernawati, I. (2020). Analisa sentimen cyberbullying di jejaring sosial twitter dengan algoritma naïve bayes. *Senamika*, 1(2), 529-538.

Maulidina, M. (2020). Analisis Sentimen Komentar Warganet Terhadap Postingan Instagram Menggunakan Metode Naive Bayes Classifier dan TF-IDF (Studi Kasus: Instagram Gubernur Jawa Barat Ridwan Kamil) (Doctoral dissertation, University of Technology Yogyakarta).

Mihalcea, R., Ignatow, G. (2016). Text Mining: A Guidebook for the Social Sciences. Amerika Serikat: SAGE Publications.

Novendri, R., Callista, A. S., Pratama, D. N., & Puspita, C. E. (2020). Sentiment analysis of YouTube movie trailer comments using naïve bayes. *Bulletin of Computer Science and Electrical Engineering*, 1(1), 26-32.

Novendri, R., Callista, A. S., Pratama, D. N., & Puspita, C. E. (2020). Sentiment analysis of YouTube movie trailer comments using naïve bayes. *Bulletin of Computer Science and Electrical Engineering*, 1(1), 26-32.

Parjono, P., & Kusumadewi, S. (2023). Pemodelan Text Mining dalam Pengkodean Penyakit Pasien Berdasar Kode ICD 10. *Jurnal Nasional Teknologi dan Sistem Informasi*, 9(2), 200-207.

- Raschka, S., Patterson, J., & Nolet, C. (2020). Machine learning in python: Main developments and technology trends in data science, machine learning, and artificial intelligence. *Information*, 11(4), 193.
- Rosid, M. A., Fitrani, A. S., Astutik, I. R. I., Mulloh, N. I., & Gozali, H. A. (2020, June). Improving text preprocessing for student complaint document classification using sastrawi. In *IOP Conference Series: Materials Science and Engineering* (Vol. 874, No. 1, p. 012017). IOP Publishing.
- Rumra, N., & Rahayu, B. (2021). Perilaku Cyberbullying pada Remaja. *Jurnal Ilmiah Kesehatan Jiwa*, 3(1), 41-48. Retrieved from <https://jurnal.rs-amino.jatengprov.go.id/index.php/JIKJ/article/view/32>
- Sasmita, A. B., Rahayudi, B., & Muflikhah, L. (2022). Analisis Sentimen Komentar pada Media Sosial Twitter tentang PPKM Covid-19 di Indonesia dengan Metode Naïve Bayes. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer* e-ISSN, 2548, 964X.
- Suraya, S., & Sholeh, M. (2022). Designing and implementing a database for thesis data management by using the python Flask Framework. *International Journal of Engineering, Science and Information Technology*, 2(1), 9-14.
- Syarif, R. D., Herdiani, A., & Astuti, W. (2019). Identifikasi Cyberbullying Pada Komentar Instagram Menggunakan Metode *Lexicon-based* Dan *Naïve Bayes Classifier* (studi Kasus: Pemilihan Presiden Indonesia Tahun 2019). *eProceedings of Engineering*, 6(2).
- Townsend, J. T. (1971). Theoretical analysis of an alphabetic confusion matrix. *Perception & Psychophysics*, 9, 40-50.
- Wahyunita, S., Azhar, Y., & Hayatin, N. (2020). Analisa Sentimen Tweet Berbahasa Indonesia dengan Menggunakan Metode Pembobotan Hybrid TF-IDF pada Topik Transportasi Online. *Jurnal Repositor*, 2(2), 185-192.
- Wardhani, D., Astuti, R., & Saputra, D. D. (2024). Optimasi Feature Selection Text Mining: Stemming dan Stopword untuk Sentimen Analisis Aplikasi

SatuSehat. Innovative: Journal Of Social Science Research, 4(1), 7537-7548.

Whittaker, E., & Kowalski, R.M. (2015). *Cyberbullying Via Social Media. Journal of School Violence*, 14, 11 - 29.

Wijayanto A.X., Fitriani R.L, & Nurhajati L. (2019). Mencegah dan Bengatasi Bullying didunia Digital

Winahyu, J., & Suharjo, I. (2021). Aplikasi Web Analisis Sentimen Dengan Algoritma Multinomial Naïve Bayes. *KARMAPATI (Kumpulan Artikel Mahasiswa Pendidikan Teknik Informatika)*, 10(2), 206-214.

Yulita, W. (2021). Analisis sentimen terhadap opini masyarakat tentang vaksin covid-19 menggunakan algoritma naïve bayes classifier. *Jurnal Data Mining dan Sistem Informasi*, 2(2), 1-9.



UNIVERSITAS ISLAM  
**RADEN RAHMAT**