

## **BAB V**

### **KESIMPULAN DAN SARAN**

#### **5.1 Kesimpulan**

Berdasarkan hasil penelitian analisis sentimen terhadap ulasan pengguna aplikasi Brainly yang telah dipaparkan pada Bab IV, maka dapat ditarik kesimpulan sebagai berikut:

1. Pengumpulan dan pengolahan data ulasan pengguna aplikasi Brainly berhasil dilakukan melalui teknik scraping dari Google Play Store, yang kemudian diproses melalui tahapan preprocessing berupa case folding, tokenisasi, penghapusan stopword, dan stemming. Selanjutnya, data diberi label sentimen secara otomatis menggunakan model IndoBERT sehingga menghasilkan dataset berlabel yang dapat digunakan dalam proses analisis sentimen, meskipun pelabelan bersifat pseudo-label dan bergantung pada performa model pelabel otomatis.
2. Metode Naïve Bayes dan Support Vector Machine (SVM) berhasil diterapkan untuk mengklasifikasikan ulasan pengguna ke dalam tiga kelas sentimen, yaitu positif, negatif, dan netral, dengan representasi fitur menggunakan TF-IDF. Hal ini menunjukkan bahwa kedua algoritma secara teknis dapat digunakan untuk melakukan klasifikasi sentimen pada data ulasan aplikasi Brainly.
3. Hasil evaluasi kinerja model menunjukkan perbedaan performa yang signifikan antara kedua algoritma. Naïve Bayes memperoleh akurasi sebesar 74%, namun hasil confusion matrix dan nilai recall kelas negatif (0,04) serta netral (0,00) menunjukkan bahwa model sangat bias terhadap kelas mayoritas dan gagal mengenali kelas minoritas. Dengan demikian, nilai akurasi yang diperoleh tidak merepresentasikan kemampuan model secara menyeluruh.
4. Support Vector Machine (SVM) menunjukkan performa yang lebih stabil dan seimbang dibandingkan Naïve Bayes, dengan akurasi sebesar 79% serta peningkatan recall pada kelas negatif hingga 0,58. Meskipun performa pada kelas netral masih terbatas (recall 0,13), SVM terbukti lebih adaptif dalam menangani data tidak seimbang, yang tercermin dari nilai F1-Score dan

Macro Average yang lebih baik dibandingkan Naïve Bayes. Kelebihan dan kekurangan masing-masing metode dapat diidentifikasi secara jelas berdasarkan hasil pengujian. Naïve Bayes unggul dari sisi kesederhanaan dan efisiensi komputasi, namun memiliki keterbatasan serius dalam kondisi distribusi data yang tidak seimbang. Sebaliknya, SVM lebih efektif dalam memisahkan kelas sentimen pada data teks berdimensi tinggi berbasis TF-IDF, sehingga lebih mampu mengenali kelas minoritas meskipun tanpa penerapan teknik penanganan imbalance seperti oversampling atau class weighting.

## 5.2 Saran

Berdasarkan hasil penelitian dan keterbatasan yang telah diidentifikasi, maka beberapa saran yang bersifat teknis dan aplikatif dapat diajukan untuk penelitian selanjutnya sebagai berikut:

1. Proses pelabelan sentimen pada penelitian selanjutnya disarankan untuk tidak sepenuhnya bergantung pada pelabelan otomatis oleh IndoBERT. Validasi manual dapat dilakukan pada sebagian data, khususnya pada kelas minoritas (negatif dan netral), untuk mengevaluasi konsistensi pseudo-label serta mengurangi propagasi kesalahan pelabelan ke tahap pelatihan model klasifikasi.
2. Untuk mengatasi permasalahan ketidakseimbangan data, penelitian lanjutan disarankan menerapkan teknik penanganan data imbalance secara sistematis, seperti Synthetic Minority Over-sampling Technique (SMOTE) atau pemberian bobot kelas (class weighting) pada algoritma Naïve Bayes dan SVM. Penerapan teknik ini diharapkan dapat meningkatkan kemampuan model dalam mengenali kelas minoritas tanpa mengorbankan performa kelas mayoritas.
3. Dari sisi metodologi pelabelan, penelitian selanjutnya dapat mengembangkan pendekatan dengan melakukan fine-tuning model IndoBERT menggunakan data ulasan yang telah divalidasi secara manual. Dengan demikian, model IndoBERT tidak hanya berfungsi sebagai alat pelabelan awal, tetapi juga sebagai model yang lebih adaptif terhadap konteks domain aplikasi Brainly, sehingga kualitas pseudo-label yang dihasilkan dapat meningkat.

4. Selain itu, penelitian berikutnya dapat menguji pendekatan klasifikasi end-to-end berbasis deep learning, di mana model transformer seperti BERT digunakan langsung untuk proses klasifikasi sentimen tanpa pemisahan tahap ekstraksi fitur TF-IDF. Pendekatan ini berpotensi menangkap konteks semantik kalimat secara lebih utuh, terutama pada ulasan yang bersifat ambigu atau kompleks.
5. Penelitian lanjutan juga disarankan untuk memperluas jumlah dan variasi data ulasan, termasuk ulasan dengan bahasa tidak formal, singkatan, atau campuran bahasa, agar model memiliki kemampuan generalisasi yang lebih baik. Evaluasi performa tetap perlu dilakukan menggunakan metrik komprehensif, namun hasilnya sebaiknya dianalisis dengan mempertimbangkan kualitas pelabelan sebagai faktor utama yang memengaruhi validitas kesimpulan.

## DAFTAR PUSTAKA

- Abidin, M. A. Z. (2025). Evaluasi sentimen ulasan pengguna CGV Cinemas Indonesia menggunakan metode Naïve Bayes dan Support Vector Machine. *Indonesian Journal on Software Engineering (IJSE)*, 11(1), 44–53.
- Al Lutfani, T. K., Astuti, R., Prihartono, W., & Hamonangan, R. (2025). Penerapan Naïve Bayes untuk analisis sentimen pada ulasan pelanggan di Lazada: studi kasus toko mawar collection. *Jurnal Informatika dan Teknik Elektro Terapan*, 13(2).
- Alfarizi, M. R. S., Al-farish, M. Z., Taufiqurrahman, M., Ardiansah, G., & Elgar, M. (2023). Penggunaan python sebagai bahasa pemrograman untuk Machine Learning dan deep learning. *Karimah Tauhid*, 2(1), 1–6.
- Arifqi, T., Suarna, N., & Prihartono, W. (2024). Penggunaan Naïve Bayes dalam menganalisis sentimen ulasan aplikasi McDonald's di Indonesia. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(2), 1949–1956.  
<https://doi.org/10.36040/jati.v8i2.8740>
- Ayuningtyas, N., & Yustanti, W. (2024). Semi-supervised learning pada pelabelan dalam klasifikasi multi-label data teks. *Journal of Informatics and Computer Science (JINACS)*, 6(01), 240–248.
- Chandradev, V., Suarjaya, I. M. A. D., & Bayupati, I. P. A. (2023). Analisis sentimen review hotel menggunakan metode deep learning BERT. *Jurnal Buana Informatika*, 14(02), 107–116.
- Choirunnisa, M., Hidayat, N., & Santoso, E. (2021). Implementasi metode Support Vector Machine dengan query expansion pada klasifikasi review di situs Traveloka. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 5(5), 1860–1865.

- Fahmi, M. N. (2023). Implementasi mechine learning menggunakan Python Library: Scikit-Learn (supervised dan unsupervised learning). *Sains Data Jurnal Studi Matematika dan Teknologi*, 1(2), 87–96.
- Harjanta, A. T. J. (2015). Preprocessing text untuk meminimalisir kata yang tidak berarti dalam proses text mining. *Jurnal Informatika Upgris*, 1(1).
- Hasanah, W. P. (2024). *Usulan perbaikan kualitas aplikasi Reddoorz berdasarkan ulasan pengguna pada Google Playstore menggunakan text mining* [Thesis, Universitas Islam Indonesia].
- Isnaeni, A. N. (2023). *Analisis higher order thinking skill siswa terhadap pengguna Brainly berdasarkan gaya belajar pada kategori reading-writing* [Undergraduate, UNIVERSITAS ISLAM SULTAN AGUNG].
- Kusuma, M. T. A., & Muharom, F. (2025). Transformasi peran pendidik dan tren pembelajaran digital di era teknologi. *Indonesian Journal of Community Engagement*, 1(2), 84–97.
- Mola, S. A. S., Djawa, S. N. R., & Mauko, A. Y. (2025). *Text mining: analisis sentimen dengan Naïve Bayes*. Kaizen Media Publishing.
- Nugroho, R. A. (2021). Pengaruh penggunaan Brainly terhadap hasil belajar siswa. *Jurnal Epistema*, 2(2), 76–92.
- Nurwanda, N., Suarna, N., & Prihartono, W. (2024). Penerapan NLP (natural language processing) dalam analisis sentimen pengguna Telegram di Playstore. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(2), 1841–1846.
- Octarina, S., Puspita, F. M., Yuliza, E., & Indrawati, I. (2025). Pendampingan penggunaan Google Colab pada pembelajaran python dan Machine Learning bagi dosen matematika di Palembang. *Jurnal Pepadu*, 6(1), 56–66.
- Prasetyo, S. D., Hilabi, S. S., & Nurapriani, F. (2023). Analisis sentimen relokasi ibukota nusantara menggunakan algoritma Naïve Bayes dan KNN. *Jurnal KomtekInfo*, 1–7.

Putra, R. R., Putri, N. A., & Putra, A. D. (2024). *Teknik cosine similarity dan tf-idf dalam analisis data*. Serasi Media Teknologi.

Putri, N. A. R., & Ardiansyah, A. (2023). Analisis sentimen terhadap kemajuan kecerdasan buatan di indonesia menggunakan BERT dan Roberta. *JURNAL SAINS DAN INFORMATIKA*, 9(2), 136–145.

Ramadhani, S. H., & Wahyudin, M. I. (2022). Analisis sentimen terhadap vaksinasi astra zeneca pada Twitter menggunakan metode NAÏVE BAYES dan K-NN. *Jurnal JTIK (Jurnal Teknologi Informasi Dan Komunikasi)*, 6(4), 526–534.

Wihardjo, E. (2025). Manajemen data dengan software modern. *Manajemen Data dengan Software Modern*, 37.

Yusepa, W. Y., Kamilawati, A., Agesti, P. T., & Sanjaya, V. F. (2024). Peran competitive advantage dalam menarik konsumen di pasar lokal (studi pada Topshop Bandar Lampung). *Al-A'mal: Jurnal Manajemen Bisnis Syariah*, 1(2), 29–40.