

Pemanfaatan Bi-LSTM dan GloVe dengan SMOTE untuk Menganalisis Sentimen Pengguna Bank Digital

Ni Komang Ayu Juliana^{a1}, Made Agung Raharja^{a2}, Ngurah Agus Sanjaya ER^{a3}, Luh Arida Ayu Rahning Putri^{a4}

Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Udayana, Bali
Jln. Raya Kampus UNUD, Bukit Jimbaran, Kuta Selatan, Badung, 08261, Bali, Indonesia
¹juliana.2208561046@student.unud.ac.id
²made.agung@unud.ac.id
³agus_sanjaya@unud.ac.id
⁴rahningputri@unud.ac.id

Abstract

The rapid growth of digital banking in Indonesia has generated a large volume of user reviews on platforms such as Google Play Store, which contain valuable feedback regarding service quality. This study aims to perform sentiment analysis on digital banking user reviews, specifically the Bank Jago application, using the Bidirectional Long Short-Term Memory (Bi-LSTM) method combined with Global Vectors for Word Representation (GloVe) Word Embedding and Synthetic Minority Over-sampling Technique (SMOTE). A total of 5,590 reviews were collected through web scraping from February 1 to September 29, 2023. Text preprocessing included cleaning, case folding, tokenization, normalization, stopword removal, and stemming. SMOTE was applied to balance the training data, resulting in a 50:50 class distribution. Hyperparameter tuning using Grid Search identified the optimal configuration of 32 hidden units and 0.0 dropout rate. The model achieved 89.09% validation accuracy and an average validation accuracy of 89.13% using 5-Fold Cross Validation. Evaluation on the test set produced an accuracy of 88.55%, demonstrating that the proposed approach effectively classifies sentiment in digital banking reviews.

Keywords: Bi-LSTM, Digital Banking, GloVe Word Embedding, Hyperparameter Tuning, K-Fold Cross Validation, Natural Language Processing, Sentiment Analysis, SMOTE, Text Preprocessing

1. Pendahuluan

Pesatnya pertumbuhan perbankan digital di Indonesia mendorong meningkatnya volume ulasan pengguna di platform seperti Google Play Store. Bank Jago, sebagai salah satu bank digital terkemuka di Indonesia, telah memiliki lebih dari 10 juta pengguna dan 225 ribu ulasan yang tersebar di platform tersebut. Ulasan-ulasan ini mengandung informasi berharga mengenai kepuasan pengguna dan kualitas layanan, namun memiliki sifat tidak terstruktur dan berjumlah sangat besar sehingga tidak memungkinkan untuk dianalisis secara manual. Oleh karena itu, analisis sentimen sebagai salah satu cabang dari *Natural Language Processing* (NLP) menjadi solusi yang efektif untuk mengklasifikasikan opini pengguna secara otomatis ke dalam kategori positif atau negatif [1].

Beberapa penelitian terdahulu telah mengeksplorasi berbagai pendekatan dalam analisis sentimen ulasan aplikasi mobile. Terdapat penelitian yang menggunakan Bi-LSTM pada ulasan aplikasi Grab di Google Play Store dan memperoleh akurasi terbaik sebesar 91%, mengungguli metode *Multinomial Naïve Bayes*, *Logistic Regression*, dan *Linear SVC* [2]. Lalu terdapat juga penelitian yang mengkombinasikan CNN dengan *GloVe Word Embedding* dan SMOTE untuk analisis sentimen di Twitter, menghasilkan akurasi 95,56% dan membuktikan bahwa perluasan fitur GloVe dikombinasikan SMOTE memberikan hasil terbaik [3]. Selain itu, terdapat juga menerapkan *Random Forest* dengan SMOTE dan *Bag of Words* pada ulasan aplikasi PLN Mobile, memperoleh akurasi 81% dan membuktikan bahwa SMOTE efektif meningkatkan performa pada dataset tidak seimbang [4]. Sementara itu, terdapat penelitian yang mengkombinasikan Bi-LSTM dengan CNN dan GloVe berbasis

TF-IDF untuk analisis sentimen Twitter dan mencapai nilai evaluasi hingga 0,993, namun tidak menangani permasalahan ketidakseimbangan kelas dalam datasetnya [5].

Meskipun penelitian-penelitian tersebut menunjukkan hasil yang menjanjikan, terdapat beberapa kekurangan yang perlu diatasi. Pertama, sebagian besar penelitian yang menggunakan Bi-LSTM belum mengkombinasikannya dengan *GloVe Word Embedding* yang dilatih secara khusus pada korpus domain perbankan digital, sehingga representasi semantik kata belum optimal untuk konteks tersebut. Kedua, penelitian yang menggunakan GloVe maupun Bi-LSTM umumnya belum menangani masalah ketidakseimbangan kelas secara bersamaan, padahal ulasan aplikasi di Google Play Store cenderung tidak seimbang antara kelas positif dan negatif [6]. Ketiga, sebagian besar penelitian belum melakukan tuning hyperparameter secara sistematis menggunakan *Grid Search*, sehingga konfigurasi model yang digunakan belum tentu optimal.

Berdasarkan kesenjangan tersebut, penelitian ini mengusulkan integrasi tiga komponen sekaligus, yaitu Bi-LSTM sebagai model klasifikasi sekuensial dua arah, *GloVe Word Embedding* yang dilatih secara custom pada korpus ulasan perbankan digital untuk menghasilkan representasi semantik yang lebih relevan, dan SMOTE untuk mengatasi ketidakseimbangan kelas pada data *training*. Kebaruan (novelty) penelitian ini terletak pada penerapan custom GloVe Word Embedding yang dilatih langsung dari data domain perbankan digital berbahasa Indonesia, sehingga mampu menangkap hubungan semantik kata yang spesifik terhadap domain tersebut, dikombinasikan dengan SMOTE berbasis *word mixing* pada level teks sebelum proses *embedding*. Selain itu, dilakukan *tuning hyperparameter* secara sistematis menggunakan *Grid Search* untuk memperoleh konfigurasi Bi-LSTM yang optimal. Pendekatan ini dievaluasi menggunakan akurasi, presisi, *recall*, *F1-score*, dan *5-Fold Cross Validation* pada dataset ulasan aplikasi Bank Digital Jago yang berjumlah 5.590 ulasan.

2. Metode Penelitian

Penelitian ini melakukan analisis sentimen terhadap ulasan pengguna aplikasi Bank Digital Jago di Google Play Store menggunakan metode Bi-LSTM dengan *GloVe Word Embedding* dan SMOTE. Langkah-langkah penelitian meliputi pengumpulan data, text preprocessing, pelabelan data, penyeimbangan kelas dengan SMOTE, representasi vektor menggunakan GloVe, pemodelan dengan Bi-LSTM, *tuning hyperparameter*, dan evaluasi model menggunakan *confusion matrix*.

2.1 Pengumpulan Data

Dataset yang digunakan diperoleh dari hasil web scraping pada Google Play Store, khususnya pada aplikasi Bank Digital Jago, menggunakan library `google_play_scraper`. Data disimpan dalam format `.csv` dan terdiri dari 5.590 ulasan yang dikumpulkan pada periode 1 Februari 2023 hingga 29 September 2023. Setiap data memiliki atribut *score/rating*, tanggal ulasan (*at*), dan isi ulasan (*content*) seperti ditunjukkan pada Tabel 1.

Tabel 1. Dataset Ulasan Bank Jago

Score	At	Content
5	2023-09-29 16:18:45	Terima kasih bank jago
2	2023-09-29 15:34:14	Pas verivikasi akhir vidio kol kok suara nya ngk keluar padahal jaringan stabil
5	2023-09-29 13:33:07	mantab lahhh nyimpen saldo dapet bunga
1	2023-09-29 02:57:00	Lah ga bisa transaksi apa apa, di atm juga ga bisa, masa scam?

2.2 Pelabelan Data

Proses pelabelan dilakukan berdasarkan rating pengguna. Ulasan dengan rating 1–2 diberi label 0 (negatif), sedangkan ulasan dengan rating 4–5 diberi label 1 (positif). Rating 3 dihapus karena dianggap netral dan tidak dapat diklasifikasikan secara tegas ke dalam salah satu kelas sentimen, sehingga dataset lebih terfokus pada sentimen yang jelas. Hasil pelabelan divalidasi oleh pakar untuk memastikan konsistensi label.

2.3 Text Preprocessing

Text preprocessing dilakukan untuk mempersiapkan data teks agar lebih terstruktur dan konsisten sebelum dianalisis. Terdapat beberapa tahapan preprocessing yang dilakukan, yaitu:

- a. *Cleaning*: penghapusan elemen yang tidak diperlukan seperti mention, hashtag, URL, karakter khusus, serta data kosong dan duplikat.
- b. *Case Folding*: konversi seluruh teks menjadi huruf kecil untuk mencegah duplikasi akibat perbedaan kapitalisasi.
- c. *Tokenization*: pemecahan teks menjadi unit kata yang lebih kecil.
- d. Normalisasi: konversi kata slang atau tidak baku ke bentuk standar menggunakan kamus slangwords.
- e. *Stopword Removal*: penghapusan kata-kata yang sering muncul tetapi tidak memiliki makna signifikan, dengan tetap mempertahankan kata-kata penting terkait sentimen dan domain perbankan.
- f. *Stemming*: pengembalian kata ke bentuk dasarnya dengan menghilangkan imbuhan menggunakan library Sastrawi.

2.4 SMOTE (*Synthetic Minority Over-sampling Technique*)

Setelah *preprocessing*, dataset dibagi menjadi 80% data *training*, 10% data validasi, dan 10% data *testing*. Untuk mengatasi ketidakseimbangan kelas pada data training, diterapkan teknik SMOTE. Berbeda dengan *random oversampling* sederhana yang hanya menduplikasi sampel secara identik sehingga rentan terhadap *overfitting*, SMOTE menciptakan sampel sintetis baru yang bervariasi melalui mekanisme interpolasi antara sampel minoritas dan tetangga terdekatnya [7].

Karena data yang diproses berupa teks mentah yang bersifat diskrit, interpolasi numerik langsung sebagaimana formula asli SMOTE tidak dapat diterapkan. Oleh karena itu, digunakan pendekatan *word mixing* sebagai realisasi tekstual dari prinsip interpolasi SMOTE. Pendekatan ini selaras dengan penelitian Wei & Zou [8] yang membuktikan bahwa augmentasi berbasis operasi kata efektif meningkatkan performa klasifikasi teks, serta didukung oleh Guo *et al* [9] yang menunjukkan bahwa pencampuran pada level token menghasilkan augmentasi yang representatif. Pemilihan pendekatan ini juga mempertimbangkan bahwa augmentasi dilakukan sebelum tahap embedding, sehingga setiap kata dalam sampel sintetis tetap berasal dari kosakata asli data pelatihan dan tidak mengintroduksi token asing yang dapat mengganggu proses vektorisasi berikutnya.

Proses *word mixing* dilakukan dengan memilih secara acak dua sampel dari kelas yang sama (*Parent 1* dan *Parent 2*), kemudian mengkombinasikan kata-kata dari keduanya dengan probabilitas 50:50 pada setiap posisi kata. Jika salah satu teks lebih pendek, kata-kata dari teks yang lebih panjang digunakan untuk posisi selanjutnya. Sebagai contoh, jika teks1 = "*aplikasi sering error tidak bisa login sangat mengecewakan*" dan teks2 = "*sering crash tutup sendiri tidak stabil buruk*", maka teks sintetis yang dihasilkan dapat berupa "*aplikasi crash error sendiri bisa stabil sangat mengecewakan*". Pemilihan *Parent 1* dan *Parent 2* selalu dibatasi dalam kelas yang sama sehingga tidak terjadi pencampuran lintas kelas.

2.5 GloVe Word Embedding

Tahap selanjutnya adalah mengubah data teks hasil SMOTE menjadi representasi vektor numerik menggunakan GloVe (*Global Vectors for Word Representation*). GloVe memodelkan hubungan statistik antar kata berdasarkan frekuensi kemunculan bersama dalam jendela konteks tertentu [10]. Dalam penelitian ini, GloVe diimplementasikan secara custom dengan parameter *embedding_dim*=200 dan *window_size*=5. Proses meliputi pembangunan vocabulary, pembangunan co-occurrence matrix dengan pembobotan berbasis jarak ($\text{weight} = 1/\text{distance}$), pelatihan model selama 50 epoch dengan learning rate 0,05, dan transformasi teks menjadi vektor 200 dimensi melalui average pooling.

2.6 Bi-LSTM (*Bidirectional Long Short-Term Memory*)

Vektor GloVe digunakan sebagai input pada model Bi-LSTM untuk klasifikasi sentimen. Bi-LSTM memproses urutan vektor kata dari dua arah secara bersamaan, yaitu *forward* (kiri ke kanan) dan *backward* (kanan ke kiri), sehingga mampu menangkap konteks kalimat secara menyeluruh. *Hidden state* akhir dari kedua arah digabungkan (*concatenate*) menjadi vektor representasi, kemudian diproyeksikan ke dense layer dan dikonversi ke probabilitas kelas menggunakan fungsi *softmax*. Adapun persamaan dari Bi-LSTM adalah sebagai berikut:

$$\text{Forward LSTM: } h_t \rightarrow = \sigma(W_f \cdot [x_t, h_{t-1} \rightarrow] + b_f) \quad (2)$$

$$\text{Backward LSTM: } h_t \leftarrow = \sigma(W_b \cdot [x_t, h_{t+1} \leftarrow] + b_b) \quad (3)$$

$$\text{Output: } h_t = [h_t \rightarrow, h_t \leftarrow] \quad (4)$$

Pelatihan dilakukan selama 20 epoch menggunakan *categorical cross-entropy loss*, *stochastic gradient descent* (SGD), dan gradient clipping untuk mencegah *exploding gradient*.

2.7 Evaluasi Model

Evaluasi model menggunakan confusion matrix yang mengukur kinerja klasifikasi berdasarkan empat komponen: True Positives (TP), False Positives (FP), True Negatives (TN), dan False Negatives (FN). Metrik yang dihitung meliputi:

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

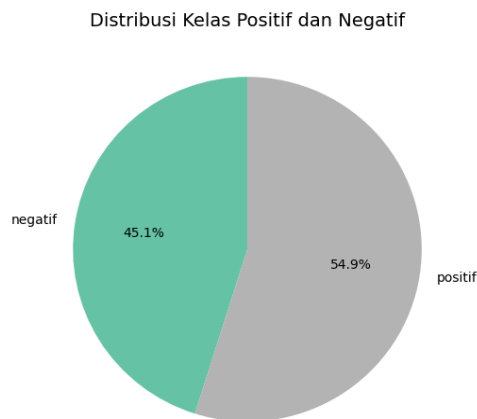
$$F1\ Score = 2 \times \frac{precision \times recall}{precision + recall} \quad (8)$$

Selain evaluasi standar, dilakukan pula 5-Fold Cross Validation pada model terbaik untuk memastikan kemampuan generalisasi model dan kestabilan performa di berbagai pembagian data.

3. Hasil dan Pembahasan

3.1 Hasil Pengumpulan Data dan Pelabelan

Dari hasil web scraping, diperoleh 5.590 ulasan pengguna aplikasi Bank Jago periode 1 Februari hingga 29 September 2023. Setelah proses pelabelan berdasarkan rating, distribusi kelas pada dataset menunjukkan adanya ketidakseimbangan: kelas positif (label 1) sebanyak 3.069 data atau 54,9%, dan kelas negatif (label 0) sebanyak 2.521 data atau 45,1%. Berikut adalah visualisasi distribusi label data yang ditunjukkan oleh Gambar 1.



Gambar 1. Distribusi Label Data

3.2 Hasil Text Preprocessing

Setelah melalui tahapan *preprocessing* yang mencakup cleaning, case folding, tokenization, normalisasi, stopword removal, dan stemming, data teks menjadi lebih bersih dan terstruktur. Tahap normalisasi menggunakan kamus slangwords berhasil mengonversi kata-kata tidak baku seperti "ngk" menjadi "tidak", "vidio" menjadi "video", dan "nyimpen" menjadi "simpan". Tahap stopword removal mempertahankan kata-kata penting terkait sentimen (seperti "tidak", "bagus", "kecewa") dan domain perbankan (seperti "transfer", "saldo", "transaksi") agar tidak ikut terhapus. Berikut adalah contoh hasil dari penerapan *text preprocessing* pada data yang digunakan yang ditunjukkan pada Tabel 2.

Tabel 2. Contoh Hasil Text Preprocessing

Data	Hasil Preprocessing
Terima kasih bank jago 🥰	['terima', 'kasih', 'bank', 'jago']
Pas verivikasi akhir vidio kol kok suara nya ngk keluar padahal jaringan stabil	['pas', 'verifikasi', 'akhir', 'video', 'call', 'suara', 'tidak', 'keluar', 'padahal', 'jaring', 'stabil']
mantab lahhh nyimpen saldo dapet bunga	['mantap', 'simpan', 'saldo', 'dapat', 'bunga']

Lah ga bisa transaksi apa apa, di atm juga ga bisa, masa scam?

['tidak', 'bisa', 'transaksi', 'atm', 'tidak', 'bisa', 'masa', 'scam']

3.3 Hasil SMOTE dan *Splitting Data*

Dataset dibagi menjadi data training (80% = 4.472 data), data validasi (10% = 559 data), dan data testing (10% = 559 data). Setelah penerapan SMOTE pada data training, distribusi kelas menjadi seimbang sempurna dengan proporsi 50:50. SMOTE berhasil membangkitkan 442 sampel sintetis untuk kelas negatif, sehingga jumlahnya meningkat dari 2.015 menjadi 2.457 sampel, setara dengan kelas positif. Total data training meningkat dari 4.472 menjadi 4.914 sampel. Distribusi kelas sebelum dan sesudah SMOTE ditunjukkan pada Tabel 3.

Tabel 3. Distribusi Kelas Sebelum dan Sesudah SMOTE

Kelas	Sebelum SMOTE	Persentase	Sesudah SMOTE	Persentase
Negatif (0)	2.015	45,1%	2.457	50%
Positif (1)	2.457	54,9%	2.457	50%
Total	4.472	100%	4.914	100%

3.4 Hasil *GloVe Word Embedding*

Proses pelatihan GloVe berhasil membangun vocabulary sebanyak 5.244 kata unik dari 75.868 kata dalam 4.914 dokumen training. Co-occurrence matrix menghasilkan 202.494 pasangan kata dengan total 2.098.088 parameter yang dioptimasi. Model menunjukkan konvergensi yang baik selama 50 epoch dengan penurunan loss sebesar 67,7%, dari 0,0415 pada epoch ke-10 menjadi 0,0134 pada epoch ke-50. Analisis word similarity menggunakan cosine similarity menghasilkan rata-rata 0,840, mengindikasikan embedding berhasil menangkap relasi semantik antar kata dengan baik.

3.5 Hasil Pemodelan Bi-LSTM dan *Hyperparameter Tuning*

Hyperparameter tuning menggunakan *Grid Search* dilakukan dengan menguji kombinasi hidden units (32, 64, 128) dan dropout rates (0,0; 0,1; 0,2), menghasilkan 9 konfigurasi. Setiap konfigurasi dilatih selama 20 epoch dengan learning rate 0,001. Hasil tuning ditunjukkan pada Tabel 4.

Tabel 4. Hasil Hyperparameter Tuning Model Bi-LSTM

Jumlah Units LSTM	Dropout Rate	Train Accuracy	Validation Accuracy	Validation Loss	Precision
32	0.0	89,56%	89,09%	0.2990	88,90%
32	0.2	89,28%	89,09%	0.3011	88,90%
128	0.1	89,70%	89,09%	0,3019	88,90%
128	0.2	89,50%	89,09%	0.3034	88,90%
128	0.0	90,03%	88,91%	0.3008	88,74%
32	0.1	89,30%	88,91%	0.2996	88,72%

64	0.0	89,46%	88,37%	0.3013	88,18%
64	0.2	88,87%	88,37%	0.3032	88,18%
64	0.1	88,83%	88,19%	0.3020	88,00%

Konfigurasi terbaik adalah hidden units 32 dan dropout rate 0,0 dengan validation accuracy 89,09% dan validation loss 0,2990. Hasil ini menunjukkan bahwa arsitektur yang lebih sederhana sudah optimal, karena representasi GloVe embeddings dan penyeimbangan data SMOTE sudah cukup untuk mencegah overfitting tanpa regularisasi tambahan. Peningkatan hidden units dari 32 menjadi 128 tidak memberikan peningkatan signifikan pada validation accuracy.

3.6 Hasil Evaluasi Model

Model terbaik hasil hyperparameter tuning kemudian dievaluasi menggunakan 5-Fold Cross Validation untuk memastikan kemampuan generalisasi. Hasil K-Fold Cross Validation ditunjukkan pada Tabel 5.

Tabel 5. Hasil 5-Fold Cross Validation

Fold	Validation Accuracy	Validation Loss	Precision
1	89.12%	0.2757	89,14%
2	90.22%	0.2458	90,28%
3	88.30%	0.2760	88,32%
4	89.12%	0.2639	89,16%
5	88.88%	0.2585	88,90%
Rata-rata	89,13%	0.2640	89,16%
Std. Deviasi	0,62%	0,0113	0,63%

Hasil K-Fold Cross Validation menunjukkan konsistensi performa yang tinggi dengan rata-rata validation accuracy 89,13% dan standar deviasi rendah 0,62%. Model terbaik berasal dari fold kedua dengan validation accuracy 90,22% dan validation loss 0,2458. Model ini kemudian dievaluasi pada test set yang terpisah dan menghasilkan test accuracy sebesar 88,55% (495 dari 559 sampel terklasifikasi dengan benar). Hasil evaluasi per kelas pada test set ditunjukkan pada Tabel 6.

Tabel 6. Hasil Evaluasi Model pada Test Set

Kelas	Precision	Recall	F1-Score	Support
Negatif (0)	86,52%	89,19%	87,83%	259
Positif (1)	90,41%	88,00%	89,19%	300
Average	88,46%	88,59%	88,51%	559

Hasil *test accuracy* ini dipengaruhi oleh beberapa faktor positif. Pertama, penggunaan GloVe yang dilatih khusus pada korpus domain perbankan digital menghasilkan representasi vektor kata yang relevan, sehingga model mampu membedakan kata-kata bernuansa positif dan negatif dalam konteks layanan perbankan dengan lebih baik. Kedua, penyeimbangan kelas melalui SMOTE *word mixing* terbukti efektif meningkatkan kemampuan model dalam mengenali kelas negatif, tercermin dari *recall* kelas negatif yang mencapai 89,19%. Namun, akurasi ini menurun pada kondisi ulasan yang mengandung ironi, negasi kompleks, atau kalimat ambigu. Contohnya, teks "Pelayanan mengecewakan sekali" diprediksi sebagai positif karena GloVe belum sepenuhnya menangkap nuansa intensifikasi negatif dari kata "sekali" dalam konteks tersebut, akibat minimnya pasangan *co-occurrence* yang relevan dalam korpus *training*.

Perbedaan *precision* antar kelas juga menarik untuk dianalisis. Kelas positif memperoleh *precision* lebih tinggi (90,41%) dibandingkan kelas negatif (86,52%), yang berarti ketika model memprediksi sentimen positif, prediksinya cenderung lebih akurat. Sebaliknya, kelas negatif memperoleh *recall* lebih tinggi (89,19% berbanding 88,00%), yang berarti model lebih sensitif dalam mendeteksi ulasan negatif yang sebenarnya, sehingga lebih sedikit ulasan negatif yang terlewat. Pola ini merupakan dampak positif dari penerapan SMOTE yang memperkuat kemampuan model dalam mengenali kelas negatif sebagai kelas minoritas asli sebelum augmentasi.

Jika dibandingkan dengan penelitian terdahulu yang relevan, hasil ini berada pada posisi yang kompetitif. Pada penelitian sebelumnya yang menggunakan Bi-LSTM pada domain yang serupa (ulasan aplikasi mobile) dan memperoleh akurasi 91%, lebih tinggi 2,45% dibandingkan penelitian ini [2]. Perbedaan ini dapat dikaitkan dengan penggunaan dataset yang lebih besar (5.000 data) dan tiga kelas sentimen, namun penelitian tersebut tidak menangani ketidakseimbangan kelas sehingga cenderung bias terhadap kelas mayoritas. Dibandingkan dengan penelitian lain yang menggunakan *Random Forest* dengan SMOTE dan BoW pada domain yang serupa dan menghasilkan akurasi 81% [4], pendekatan Bi-LSTM dengan GloVe dalam penelitian ini memberikan peningkatan akurasi sebesar 7,55%. Hal ini mengonfirmasi bahwa model *deep learning* berbasis sekuens dengan representasi semantik yang kaya lebih unggul dibandingkan model *machine learning* konvensional dengan representasi berbasis frekuensi untuk tugas analisis sentimen ulasan aplikasi.

4. Kesimpulan

Berdasarkan hasil penelitian, dapat disimpulkan bahwa analisis sentimen ulasan pengguna aplikasi Bank Digital Jago menggunakan metode Bi-LSTM yang dikombinasikan dengan GloVe Word Embedding dan SMOTE menunjukkan performa yang baik. Model terbaik yang diperoleh dari 5-Fold Cross Validation menghasilkan *test accuracy* sebesar 88,55% dengan *precision* 88,46%, *recall* 88,59%, dan F1-score 88,51%. Rata-rata *validation accuracy* dari K-Fold Cross Validation mencapai 89,13% dengan standar deviasi rendah sebesar 0,62%, mengindikasikan bahwa model memiliki kemampuan generalisasi yang baik dan konsisten di berbagai pembagian data.

Konfigurasi *hyperparameter* terbaik yang diperoleh melalui *Grid Search* adalah *hidden units* 32 dan *dropout rate* 0,0 dengan *validation accuracy* 89,09%, *test accuracy* 88,55% dan *validation loss* 0,2990. Hasil ini menunjukkan bahwa arsitektur Bi-LSTM yang lebih sederhana sudah cukup optimal karena representasi GloVe embeddings dan penyeimbangan data dengan SMOTE sudah efektif mencegah *overfitting* tanpa memerlukan regularisasi tambahan. Peningkatan performa lebih lanjut dapat dilakukan dengan mengeksplorasi model berbasis transformer seperti BERT atau IndoBERT, memperluas kamus *slang words*, serta menggunakan dataset yang lebih besar dan lebih beragam.

References

- [1] G. Darmawan, N. Gusriani, R. Ruslan, F. Agustiana, and R. A. Saputra, "Perbandingan Model Bidirectional-LSTM dan Singular Spectrum Analysis dalam Peramalan Harga Cabai," *Innovative: Journal Of Social Science Research*, vol. 3, no. 2, pp. 1504–1514, May 2023.
- [2] D. R. Alghifari, M. Edi, and L. Firmansyah, "Implementasi Bidirectional LSTM untuk Analisis Sentimen Terhadap Layanan Grab Indonesia," *Jurnal Manajemen Informatika (JAMIKA)*, vol. 12, no. 2, pp. 89–99, Sep. 2022, doi: 10.34010/jamika.v12i2.7764.
- [3] Diaz Tiyasya Putra and Erwin Budi Setiawan, "Sentiment Analysis on Social Media with Glove Using Combination CNN and RoBERTa," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 7, no. 3, pp. 457–563, Jun. 2023, doi: 10.29207/resti.v7i3.4892.
- [4] M. R. F. Rahmatullah, P. N. Andono, Affandy, and M. A. Soeleman, "Improving Random Forest Performance for Sentiment Analysis on Unbalanced Data Using SMOTE and BoW Integration: PLN Mobile Application Case Study," *Scientific Journal of Informatics*, vol. 12, no. 1, pp. 1–10, Apr. 2025,

doi: 10.15294/sji.v12i1.19295.

- [5] M. Kamyab, G. Liu, and M. Adjeisah, "Attention-Based CNN and Bi-LSTM Model Based on TF-IDF and GloVe Word Embedding for Sentiment Analysis," *Applied Sciences*, vol. 11, no. 23, p. 11255, Nov. 2021, doi: 10.3390/app112311255.
- [6] E. Sutoyo and M. A. Fadlurrahman, "Penerapan SMOTE untuk Mengatasi Imbalance Class dalam Klasifikasi Television Advertisement Performance Rating Menggunakan Artificial Neural Network," *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, vol. 6, no. 3, pp. 379–385, Dec. 2020, doi: 10.26418/jp.v6i3.42896.
- [7] A. Nurhopipah and C. Magnolia, "PERBANDINGAN METODE RESAMPLING PADA IMBALANCED DATASET UNTUK KLASIFIKASI KOMENTAR PROGRAM MBKM," *Jurnal Publikasi Ilmu Komputer dan Multimedia*, vol. 2, no. 1, pp. 9–22, Jan. 2023, doi: 10.55606/jupikom.v2i1.862.
- [8] J. Wei and K. Zou, "EDA: Easy Data Augmentation Techniques for Boosting Performance on Text Classification Tasks," in *Proc. 2019 Conf. Empirical Methods Natural Lang. Process. 9th Int. Joint Conf. Natural Lang. Process. (EMNLP-IJCNLP)*, Hong Kong, China, 2019, pp. 6382–6388, doi: 10.18653/v1/D19-1670.
- [9] H. Guo, Y. Mao, and R. Zhang, "Augmenting Data with Mixup for Sentence Classification: An Empirical Study," *arXiv preprint arXiv:1905.08941*, 2019. [Online]. Available: <https://arxiv.org/abs/1905.08941>
- [10] S. AI, "Sentiment Analysis — using LSTM & GloVe Embeddings," Medium. Accessed: Mar. 29, 2026. [Online]. Available: <https://medium.com/@skillcate/sentiment-classification-using-neural-networks-a-complete-guide-1798aaf357cd>
- [11] N. L. K. I. Pramesti, M. A. Raharja, N. A. Sanjaya, and I. G. A. Wibawa, "Analisis Sentimen Ulasan Aplikasi Solusi Kota Cerdas Menggunakan Algoritma Naïve Bayes dan Support Vector Machine (SVM) dengan Seleksi Fitur Chi-Square," (*JELIKU*) *Jurnal Elektronik Ilmu Komputer Udayana*, vol. 13, no. 1, pp. 1045–1058, May 2024.
- [12] K. J. Pratama and I. N. Fajri, "Sistem Informasi dan Klasifikasi Limbah Makanan Berbasis Website dengan Menggunakan Metode CNN," *JuTI (Jurnal Teknologi Informasi)*, vol. 3, no. 2, p. 123, Feb. 2025, doi: 10.26798/juti.v3i2.1854.
- [13] I. M. S. A. Palguna, N. A. S. Er, I. M. Widiartha, and M. A. Raharja, "Twitter Sentiment Analysis Regarding the Influence of Political Figures Using the Distance – Weighted K – Nearest Neighbor Method," *Jurnal Elektronik Ilmu Komputer Udayana*, vol. 14, no. 2, pp. 233–240, Nov. 2025.