



JOURNAL CERITA:

**Creative Education of Research in Information
Technology and Artificial Informatics**

Vol. 11 No. 1 (2025) 120-128

e-ISSN: 2655 - 2574

Prediksi Transaksi Minat Pembelian Online Menggunakan Kombinasi CNN Conv1D dan BiLSTM

Maimi Herawati^{*1}, Kusrini²

^{1,2}Program Pascasarjana, Universitas Amikom Yogyakarta, Indonesia

Email: ^{*1}maimiherawati@students.amikom.ac.id; ²kusrini@amikom.ac.id

Herawati, M., & Kusrini, K. (2025). Prediksi Transaksi Minat Pembelian Online Menggunakan Kombinasi *Convolutional Neural Network* (Conv1D) dan *Bidirectional Long Short-Term Memory* (BiLSTM). *Journal Cerita: Creative Education of Research in Information Technology and Artificial Informatics*, 11(1), 120-128

DOI: <https://doi.org/10.33050/cerita.v11i1.3702>

ABSTRAK

Perkembangan teknologi informasi telah mengubah cara konsumen melakukan aktivitas belanja, khususnya melalui platform e-commerce. Belanja online menjadi tren utama karena kemudahan akses dan peningkatan penetrasi internet. Memahami niat pembelian online menjadi hal krusial bagi pelaku bisnis untuk merancang strategi pemasaran yang lebih efektif. Niat pembelian konsumen dipengaruhi oleh berbagai faktor seperti kualitas produk, harga, ulasan pelanggan, dan kemudahan penggunaan platform. Namun, tantangan utama dalam memprediksi niat pembelian adalah volume data yang sangat besar dan kompleks. SMOTE digunakan untuk mengatasi ketidakseimbangan data. Penelitian ini bertujuan untuk mengembangkan model deep learning berbasis kombinasi CNN(Conv1D) dan BiLSTM untuk memprediksi niat pembelian konsumen dengan akurasi tinggi. Fokus penelitian mencakup analisis akurasi model serta efektivitas algoritma dalam mengatasi ketidakseimbangan data. Dari hasil penelitian model CNN(Conv1D) + BiLSTM mencapai akurasi 97% dengan hasil evaluasi yang seimbang, meski recall kelas True (96%) masih lebih rendah dibandingkan kelas False (95%). Optimasi lebih lanjut diperlukan untuk meningkatkan performa model.

Kata kunci: BiLSTM, CNN(Conv1D), Ketidakseimbangan data, Minat Pembelian, SMOTE

ABSTRACT

The rapid development of information technology has transformed consumer shopping behavior, particularly through e-commerce platforms. Online shopping has become a primary trend due to its convenience and the growing penetration of the internet. Understanding online purchase intention is therefore crucial for businesses in devising effective marketing strategies. Purchase intention is influenced by factors such as product quality, price, customer reviews, and platform usability. However, predicting purchase intention poses a significant challenge due to the large and complex nature of consumer data. Smote used for imbalance data. This study aims to combine CNN (Conv1D) and BiLSTM for high-accuracy purchase intention prediction. The research focuses on analyzing model accuracy and the effectiveness of the algorithms in handling imbalanced data. The results indicate that the combined CNN(Conv1D) + BiLSTM model achieves 97% accuracy with balanced evaluation metrics, although the True class recall (96%) is slightly lower than that of the False class (95%). Further optimization is needed to enhance overall model performance.

Keywords: BiLSTM, CNN(Conv1D), Imbalance data, Purchase Intention, SMOTE

I. PENDAHULUAN

Perkembangan teknologi informasi telah mengubah berbagai aspek kehidupan, termasuk cara konsumen melakukan aktivitas belanja. E-Commerce sebagai salah satu sektor yang tumbuh pesat, memberikan kemudahan bagi konsumen untuk mengakses berbagai produk dan layanan hanya melalui perangkat digital. Popularitas *platform* belanja *online* telah meningkat pesat seiring berjalannya waktu. Penetrasi internet yang cepat telah mengubah aktivitas bisnis secara masif dan bisnis diaktifkan melalui internet telah menjadi fenomena baru. Memahami perilaku belanja online pada platform e-commerce menjadi salah satu hal penting bagi pelaku bisnis untuk meningkatkan penjualan dan pengalaman pelanggan (Wang, 2023). Belanja online adalah salah satu aspek yang membuat pelanggan di era saat ini menjadi ketagihan. Dalam konteks ini, memahami perilaku konsumen, khususnya niat atau minat pembelian (*purchasing intention*) menjadi hal yang sangat penting bagi pelaku bisnis untuk meningkatkan penjualan dan pengalaman pelanggan (Dharmesti et al., 2019).

Niat pembelian online adalah konsep yang merujuk pada kecenderungan atau intensi konsumen untuk membeli produk secara daring. Hal ini dipengaruhi oleh berbagai faktor seperti kualitas produk, harga, ulasan pelanggan, dan kemudahan penggunaan platform (Yuliana et al., 2023). Memahami niat pembelian ini dapat membantu perusahaan e-commerce dalam mengembangkan strategi pemasaran yang lebih efektif dan meningkatkan konversi. Namun, menganalisis niat pembelian secara manual menjadi tantangan tersendiri mengingat

banyaknya data yang tersedia dari aktivitas pengguna di platform e-commerce (Lin & Shen, 2023). Minat pembelian merujuk pada kemungkinan atau kecenderungan seorang konsumen untuk melakukan pembelian produk atau jasa melalui platform e-commerce (Wen et al., 2023). Niat ini dipengaruhi oleh berbagai faktor seperti persepsi terhadap kualitas produk, harga, ulasan pelanggan, dan kenyamanan platform. Memahami niat pembelian ini menjadi krusial bagi perusahaan e-commerce untuk meningkatkan strategi pemasaran dan meningkatkan konversi penjualan (Grimaldo et al., 2023). Pada era digital saat ini, volume data yang dihasilkan dari aktivitas online konsumen sangat besar. Setiap interaksi pengguna dengan platform e-commerce, seperti melihat produk, menambahkan produk ke keranjang, atau membaca ulasan, menghasilkan data yang dapat digunakan untuk memahami perilaku konsumen (Le-Hoang, 2020). Namun, tantangan utama adalah bagaimana memanfaatkan data ini secara efektif untuk memprediksi niat pembelian konsumen secara akurat. Penelitian terkait prediksi minat dan transaksi saat ini terus dikembangkan baik dalam segi pengembangan model seperti *deep learning* dan model konvensional ataupun tren issue terkait minat dan transaksi dari berbagai aspek (Yoke Cheng et al., 2019). *Deep learning*, sebagai salah satu cabang dari machine learning, menawarkan pendekatan yang sangat efektif dalam memproses data dalam jumlah besar dan kompleks.

A. TINJAUAN PUSTAKA

Beberapa penelitian telah dilakukan sebelumnya untuk meningkatkan akurasi, penelitian tersebut antara lain:

Penelitian mengenai prediksi niat pembelian pelanggan online menjadi topik yang semakin relevan di tengah pesatnya pertumbuhan *e-commerce*. Tanvir dkk membangun model prediksi berbasis *real-time* yang mampu mendeteksi niat pembelian pelanggan selama sesi kunjungan berlangsung. Penelitian ini menggunakan algoritma XGBoost dengan teknik *oversampling* untuk mengatasi ketidakseimbangan data (Abdullah-All-Tanvir et al., 2023). Karim Baati mengusulkan sistem prediksi perilaku belanja secara *real-time* untuk mencegah pelanggan meninggalkan situs tanpa bertransaksi. Penelitian ini menggunakan algoritma Random Forest, Naïve Bayes, dan C4.5 Decision Tree dengan metode *oversampling* untuk mengatasi ketidakseimbangan data (Baati & Mohsil, 2020). Penelitian yang dilakukan oleh Arif Purnama berfokus pada pengujian algoritma Random Forest menggunakan pembagian dataset 60% untuk pelatihan dan 40% untuk pengujian (Purnama et al., 2021). Yadav mengevaluasi kinerja berbagai algoritma *machine learning*, seperti *Logistic Regression*, *Naïve Bayes*, *Decision Tree*, *Random Forest*, dan *Gradient Boosting*, menggunakan data *clickstream* pengguna (Yadav & Student, 2023). Membandingkan algoritma *machine learning* konvensional seperti *Decision Tree*, *Logistic Regression*, dan *Random Forest* dengan *Deep Neural Network* (DNN) dilakukan oleh Haris untuk menguji akurasi (Muhammad, 2022).

Dengan demikian, penelitian lebih lanjut tentang penggunaan *deep learning* dalam memprediksi niat pembelian online dapat memberikan kontribusi yang signifikan bagi industri *e-commerce*. Model prediksi yang lebih akurat tidak hanya akan meningkatkan pengalaman pengguna, tetapi juga membantu perusahaan dalam mengambil keputusan yang lebih baik terkait strategi pemasaran, penentuan harga, dan penawaran produk. Penelitian ini bertujuan untuk mengembangkan model *deep learning* yang dapat digunakan untuk memprediksi niat pembelian konsumen dengan akurasi tinggi, yang pada akhirnya dapat meningkatkan konversi penjualan dan loyalitas pelanggan.

Adapun fokus penelitian ini adalah:

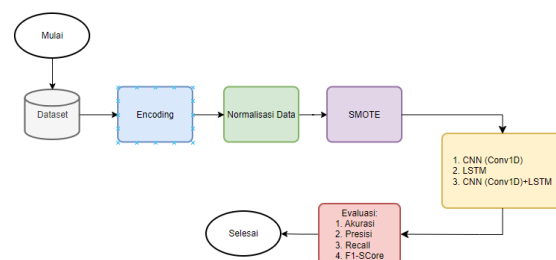
1. Bagaimana penerapan kombinasi CNN (Conv1D) dan BiLSTM pada data tabular transaksi pembelian

2. Bagaimana performa kombinasi CNN (Conv1D) dan BiLSTM .

II. METODE PENELITIAN

A. Alur Penelitian

Alur proses analisis data menggunakan pendekatan *deep learning* seperti ditunjukkan pada Gambar 1. Proses dimulai dengan pemuatan dataset sebagai sumber data utama yang akan digunakan dalam analisis. Selanjutnya, dilakukan tahap *encoding* untuk mengonversi data non-numerik menjadi data numerik yang dapat diolah oleh model pembelajaran mesin. Setelah itu, data melalui proses normalisasi agar nilai setiap fitur berada dalam rentang tertentu, biasanya antara 0 dan 1, untuk menghindari dominasi fitur tertentu yang memiliki nilai besar. Proses berikutnya adalah penerapan teknik *Synthetic Minority Over-sampling Technique* (SMOTE) untuk menyeimbangkan jumlah sampel antara kelas mayoritas dan minoritas dengan mensintesis data baru berdasarkan tetangga terdekat, sehingga mengatasi masalah ketidakseimbangan data yang dapat memengaruhi kinerja model. Data yang telah seimbang kemudian digunakan dalam tiga jenis arsitektur model pembelajaran mesin, yaitu CNN(Conv1D) untuk menangkap pola spasial dan lokal, *Bidirectional Long Short-Term Memory* (BiLSTM) untuk mengenali pola temporal pada data sekuensial, serta kombinasi CNN(Conv1D)- BiLSTM yang menggabungkan kemampuan ekstraksi fitur spasial CNN(Conv1D) dan analisis temporal BiLSTM. Setelah proses pelatihan, model dievaluasi menggunakan beberapa metrik, yaitu akurasi, presisi, *recall*, dan *F1-score*, yang bertujuan untuk mengukur kinerja model secara komprehensif dalam klasifikasi data. Akhirnya, proses analisis ini selesai setelah hasil evaluasi diperoleh dan digunakan untuk menentukan model terbaik yang paling sesuai dengan permasalahan yang dianalisis.



Gambar 1. Alur Penelitian

B. Dataset

Pengumpulan data dilakukan untuk mendapatkan berbagai macam data yang diperlukan dalam penelitian. Penelitian ini menggunakan data sekunder yang diambil dari dataset publik yang sudah tersedia yaitu *Online Shoppers' Purchasing Intention* dari *UCI machine learning repository* (Torres et al., 2024). Dataset ini terdiri dari 18 fitur *independent*, 2 label dan berjumlah 12.330 observasi yang menggambarkan sesi pengguna *e-commerce*. Data yang digunakan terdiri dari *variable – variable* yang relevan dalam memprediksi niat pembelian konsumen. Tabel 1 menunjukkan dataset.

Tabel 1. *Dataset Online Shopper Purchase Intention*

No	Nama Fitur	Type Data
1.	<i>Administrative</i>	Numerik
2.	<i>Administrative_Duration</i>	Numerik
3.	<i>Informational</i>	Numerik
4.	<i>Informational_Duration</i>	Numerik
5.	<i>ProductRelated</i>	Numerik
6.	<i>ProductRelated_Duration</i>	Numerik
7.	<i>BounceRates</i>	Numerik
8.	<i>ExitRates</i>	Numerik
9.	<i>PageValues</i>	Numerik
10.	<i>SpecialDay</i>	Numerik
11.	<i>Month</i>	Kategorikal
12.	<i>OperatingSystems</i>	Kategorikal

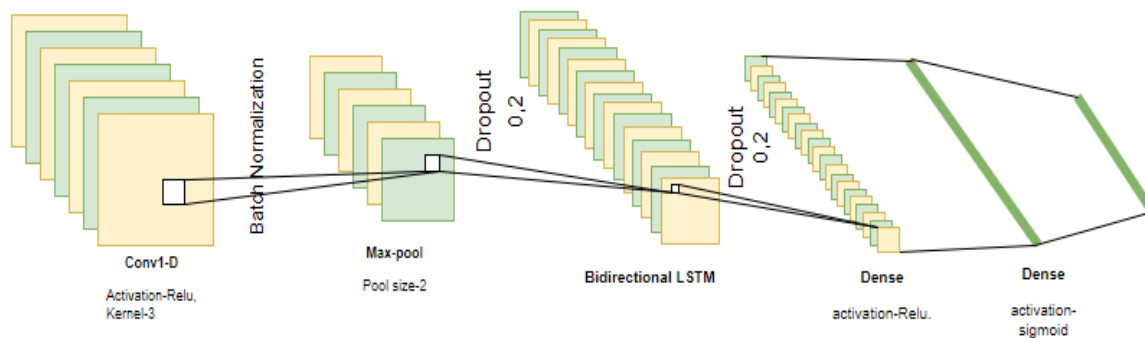
13.	<i>Browser</i>	Kategorikal
14.	<i>Region</i>	Kategorikal
15.	<i>TrafficType</i>	Kategorikal
16.	<i>VisitorType</i>	Kategorikal
17.	<i>Weekend</i>	Kategorikal
18.	<i>Revenue</i>	Kategorikal

C. Pre-Processing

Data yang digunakan tidak memiliki *missing value* sehingga pada tahapan ini dilakukan beberapa proses mulai dari data normalisasi, *data balancing*, dan *split data*. Proses normalisasi adalah proses transformasi data numerik sehingga berada dalam rentang atau skala tertentu misalnya 0 hingga 1. Hal ini bertujuan mempercepat konvergensi model yang akan diterapkan.

D. SMOTE

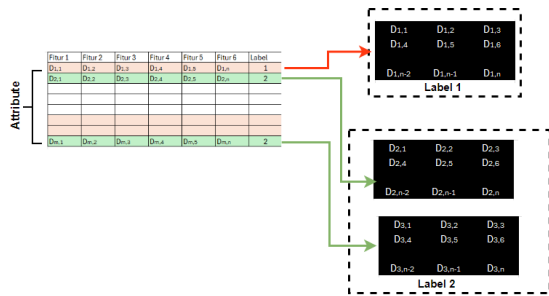
Data balancing dilakukan apabila dataset memiliki ketidakseimbangan, dengan menggunakan teknik *oversampling* yaitu (SMOTE) (Wijayanti, 2021). SMOTE menggunakan teknik *oversampling* sehingga mempunyai kelebihan yaitu tidak akan menyebabkan adanya informasi yang hilang, hal ini disebabkan tidak ada pengurangan data (Nugroho H et al., 2025). *Split Data* merupakan proses preprocessing terakhir dimana dataset akan dibagi menjadi 80 % untuk data latih dan 20 % untuk data uji.



Gambar 3. Skenario Arsitektur CNN(Conv1D) dan BiLSTM

E. CNN(Conv1D) dan BiLSTM

Setelah proses split data dilakukan, data tabular perlu dikonversi menjadi citra (Du et al., 2021). Gambar 2 menjelaskan proses transformasi data tabular menjadi representasi citra sebelum diproses oleh arsitektur CNN (Conv1D) dan BiLSTM (Mamani, 2023). Data tabular yang terdiri dari baris sebagai sampel dan kolom sebagai atribut atau fitur diubah menjadi format gambar. Setiap baris data dipecah berdasarkan kategori atau label kelas, seperti "Label 1" dan "Label 2", di mana setiap label memiliki kumpulan sub-citra yang terdiri dari nilai-nilai numerik per atribut (Kulkarni, 2024). Nilai-nilai ini kemudian disusun menjadi matriks piksel yang merepresentasikan intensitas berdasarkan nilai data numerik. Matriks ini membentuk representasi citra yang menjadi input bagi model CNN (Conv1D) dan BiLSTM.



Gambar 2. Proses transformasi data tabular menjadi citra.

Gambar 3 menunjukkan skenario arsitektur model deep learning berbasis kombinasi CNN (Conv1D)-BiLSTM yang dilakukan dalam penelitian ini. Proses dimulai dengan lapisan Conv1D yang mengekstraksi fitur lokal dari data menggunakan kernel berukuran 3 dengan fungsi aktivasi ReLU. Output dari lapisan ini dinormalisasi melalui Batch Normalization untuk menjaga kestabilan distribusi selama pelatihan.

Kemudian, Max-Pooling dengan ukuran 2 digunakan untuk mengurangi dimensi output dengan menyaring informasi penting. Selanjutnya, lapisan BiLSTM menangkap hubungan temporal dari arah maju dan mundur untuk memahami konteks data secara menyeluruh. Untuk mencegah overfitting, diterapkan Dropout dengan tingkat penghapusan neuron sebesar 0.2 di berbagai lapisan. Setelah itu, output dari LSTM diteruskan ke lapisan dense dengan fungsi aktivasi ReLU untuk memperkuat hubungan non-linear antar fitur.

Pada lapisan akhir, digunakan fungsi aktivasi sigmoid untuk menghasilkan output biner, sesuai dengan tugas klasifikasi biner. Arsitektur ini memanfaatkan kekuatan CNN (Conv1D) dalam ekstraksi fitur lokal dan BiLSTM dalam memahami hubungan jangka panjang, menjadikannya efektif untuk tugas analisis data sekuensial seperti klasifikasi teks atau data temporal (Angraini & Lhaksana, 2024; Chetana R et al., 2023).

F. Evaluasi

Confusion matrix merupakan metode yang sering digunakan untuk mengukur dan mengevaluasi performa model. *Confusion matrix* terdiri dari baris data uji yang diprediksi benar dan tidak benar suatu data oleh model klasifikasi.

$$\text{Akurasi} = \frac{(TP+TN)}{(TP+TN+FP+FN)} \quad (1)$$

$$\text{Presisi} = \frac{TP}{(TP+FP)} \quad (2)$$

$$\text{Recall} = \frac{TP}{(TP+FN)} \quad (3)$$

$$F1 - \text{Score} = 2 \times \frac{\text{presisi} \times \text{recall}}{\text{presisi} + \text{recall}} \quad (4)$$

True Positive (TP) yaitu jumlah kasus di mana model memprediksi kelas positif dan data aktualnya juga positif. *False Negative* (FN) yaitu jumlah kasus di mana model memprediksi kelas negatif, tetapi data aktualnya positif. *False Positive* (FP) yaitu jumlah kasus di mana model memprediksi kelas positif, tetapi data aktualnya negatif. *True Negative* (TN) yaitu jumlah kasus di mana model memprediksi kelas negatif dan data aktualnya juga negatif. *Confusion matrix* memberikan pandangan mendetail mengenai kesalahan prediksi model seperti akurasi, presisi, *recall*, *F-1 score*. Akurasi digunakan untuk mengukur seberapa sering model membuat prediksi yang benar, baik untuk kelas positif maupun kelas negatif, dari seluruh prediksi yang dilakukan. Presisi berfungsi mengukur seberapa banyak prediksi positif yang benar dibandingkan dengan seluruh prediksi yang dinyatakan positif oleh model. *Recall* berfungsi untuk mengukur seberapa baik model mendeteksi semua kasus positif yang sebenarnya positif di data.

III. HASIL DAN PEMBAHASAN

Gambar 4. menampilkan hasil *encoding* dari struktur dataset yang terdiri dari 12.330 baris dan 18 kolom, di mana tidak ada nilai kosong pada

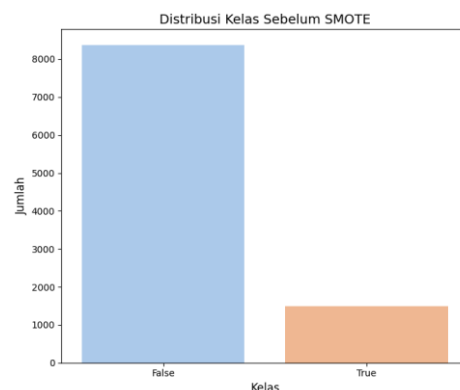
setiap kolom (semua memiliki nilai *non-null*). Dataset ini memiliki 10 kolom bertipe *integer* (int64), 7 kolom bertipe *float* (float64), dan 1 kolom bertipe *boolean* (bool), dengan total penggunaan memori sebesar 1,6 MB.

```
Data columns (total 18 columns):
# Column Non-Null Count Dtype
---
0 Administrative 12330 non-null int64
1 Administrative_Duration 12330 non-null float64
2 Informational 12330 non-null int64
3 Informational_Duration 12330 non-null float64
4 ProductRelated 12330 non-null int64
5 ProductRelated_Duration 12330 non-null float64
6 BounceRates 12330 non-null float64
7 ExitRates 12330 non-null float64
8 PageValues 12330 non-null float64
9 SpecialDay 12330 non-null float64
10 Month 12330 non-null int64
11 OperatingSystems 12330 non-null int64
12 Browser 12330 non-null int64
13 Region 12330 non-null int64
14 TrafficType 12330 non-null int64
15 VisitorType 12330 non-null int64
16 Weekend 12330 non-null int64
17 Revenue 12330 non-null bool
dtypes: bool(1), float64(7), int64(10)
memory usage: 1.6 MB
None
```

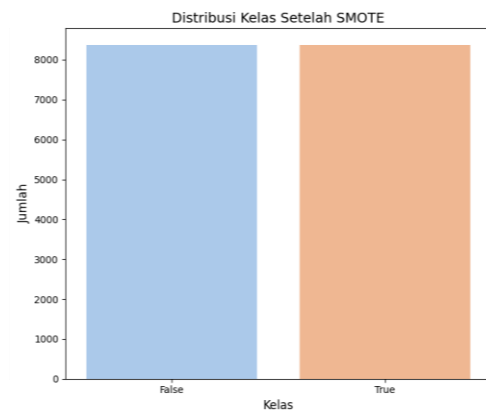
Gambar 4. Hasil *Encoding*

Administrative dan *Administrative_Duration* mencatat jumlah serta durasi kunjungan pada halaman administratif, sedangkan *Informational* dan *Informational_Duration* mencatat jumlah serta waktu yang dihabiskan pada halaman informasi. Kolom *ProductRelated* dan *ProductRelated_Duration* menunjukkan jumlah serta durasi kunjungan pada halaman produk. Fitur interaksi pengguna, seperti *BounceRates* dan *ExitRates*. Selain itu, *PageValues* menggambarkan nilai ekonomis halaman terkait konversi, sedangkan *SpecialDay* menunjukkan kedekatan waktu kunjungan dengan momen spesial, seperti promosi atau hari libur. Kolom lainnya meliputi *Month* (bulan kunjungan), *OperatingSystems* (jenis sistem operasi), *Browser* (jenis peramban), *Region* (wilayah geografis), dan *TrafficType* (sumber lalu lintas). Dataset ini juga mencatat informasi tentang *VisitorType* (jenis pengunjung, baru atau kembali), serta *Weekend* (indikator kunjungan akhir pekan). Kolom *Revenue* merupakan label target bertipe *boolean* yang menunjukkan status pembelian (*True* jika terjadi pembelian dan *False* jika tidak). Dataset ini dapat digunakan untuk menganalisis perilaku pengguna dan membangun model prediksi guna mengidentifikasi faktor-faktor yang memengaruhi keputusan pembelian dalam platform *e-commerce*.

Gambar 5 dan Gambar 6 tersebut menunjukkan grafik perbandingan kelas dalam dataset sebelum dan setelah penerapan metode SMOTE. Grafik pertama memperlihatkan distribusi kelas yang tidak seimbang, di mana jumlah sampel kelas "*False*" jauh lebih tinggi dibandingkan dengan kelas "*True*". Ketidakeimbangan ini merupakan masalah umum dalam dataset klasifikasi biner yang dapat menyebabkan model pembelajaran mesin lebih cenderung memprediksi kelas mayoritas, sehingga menurunkan performa model dalam mendeteksi kelas minoritas.



Gambar 5. Grafik perbandingan sebelum SMOTE.



Gambar 6. Grafik perbandingan setelah SMOTE

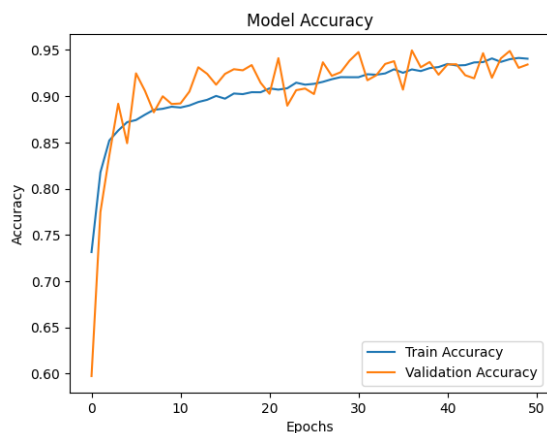
Grafik kedua menunjukkan hasil setelah penerapan SMOTE, di mana jumlah sampel antara kelas "*False*" dan "*True*" menjadi seimbang. Metode SMOTE bekerja dengan menghasilkan sampel sintesis untuk kelas minoritas berdasarkan interpolasi sampel yang ada, bukan dengan menduplikasi sampel. Hal ini membantu meningkatkan kemampuan model untuk belajar dari kedua kelas secara proporsional dan mengurangi risiko bias terhadap kelas mayoritas. Untuk melihat jumlah data dari

masing-masing kelas sebelum dan sesudah dapat dilihat pada Gambar 7.

```
Sebelum SMOTE: Revenue
False      8367
True       1497
Name: count, dtype: int64
Setelah SMOTE: Revenue
False      8367
True       8367
```

Gambar 7. Jumlah data sebelum dan setelah SMOTE.

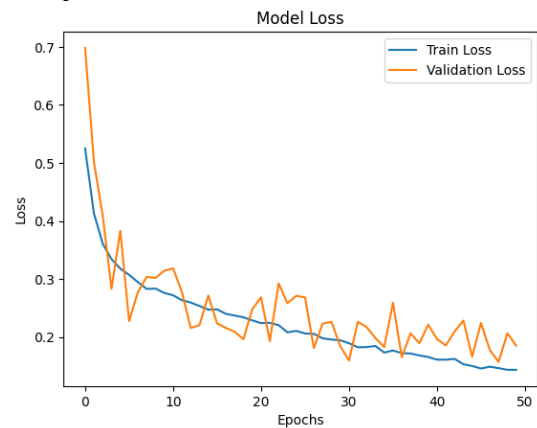
Gambar 8 menunjukkan grafik akurasi peningkatan tajam pada fase awal pelatihan, yang mengindikasikan bahwa model dengan cepat mempelajari pola penting dari data. Setelah mencapai sekitar 10 hingga 15 *epoch*, akurasi mulai mendekati titik saturasi dengan nilai berkisar antara 0,97 hingga 0,98, menunjukkan bahwa model telah mencapai tingkat akurasi yang optimal. Perbedaan yang kecil antara akurasi pelatihan dan validasi menggambarkan keseimbangan performa yang baik, yang mengindikasikan bahwa model mampu menggeneralisasi dengan efektif tanpa kecenderungan *overfitting*.



Gambar 8. Akurasi CNN (Conv1D) dan BiLSTM

Grafik loss menunjukkan penurunan yang signifikan pada epoch awal, dari nilai sekitar 0,7 menjadi mendekati 0,2 di akhir pelatihan seperti pada Gambar 9. Hal ini menunjukkan bahwa model berhasil meminimalkan kesalahan prediksi secara konsisten. Namun, fluktuasi pada loss validasi, meskipun terlihat pada beberapa titik, masih berada dalam rentang yang wajar dan tidak mengganggu pola keseluruhan. Fluktuasi ini dapat disebabkan oleh variasi data validasi yang lebih beragam, tetapi tetap memperlihatkan bahwa model tidak kehilangan kemampuan untuk mempelajari pola utama. Arsitektur CNN

(Conv1D) bertugas melakukan ekstraksi fitur spasial dari data sekuensial, seperti mendeteksi pola lokal yang relevan, sementara BiLSTM menambahkan lapisan analisis temporal dengan mempertimbangkan hubungan dua arah pada sekuens data. Sinergi kedua komponen ini memungkinkan model untuk memahami hubungan fitur secara komprehensif. Secara keseluruhan, grafik akurasi dan *loss* menunjukkan bahwa penggunaan CNN dan BiLSTM menghasilkan model yang kuat dengan akurasi tinggi dan *loss* rendah, serta mampu menjaga performa yang stabil meskipun dihadapkan pada data validasi yang berbeda dari data pelatihan.



Gambar 9. Grafik Loss CNN (Conv1D) dan BiLSTM

Hasil evaluasi model kombinasi CNN (Conv1D) dan BiLSTM menunjukkan performa yang sangat baik dalam kasus prediksi seperti Gambar 10, seperti yang terlihat dari laporan klasifikasi tersebut. Nilai *precision* untuk kelas *False* sebesar 0,96 mengindikasikan bahwa 96% prediksi untuk kelas ini adalah benar, sementara *precision* untuk kelas *True* lebih tinggi, yakni 0,98. *Recall* untuk kelas *False* mencapai 0,98, menunjukkan bahwa model mampu mengenali 98% sampel negatif dengan benar, sedangkan untuk kelas *True*, *recall* berada di angka 0,95, mengindikasikan 95% sampel positif berhasil diidentifikasi secara akurat. *F1-score*, yang merupakan metrik gabungan antara *precision* dan *recall*, menunjukkan nilai 0,97 untuk kedua kelas, menandakan keseimbangan performa model dalam mengenali kelas positif dan negatif. Jumlah sampel (*support*) untuk masing-masing kelas adalah 8367, menghasilkan total sampel evaluasi sebesar 16.734. Akurasi keseluruhan model mencapai 97%, yang menunjukkan kemampuan model untuk mengklasifikasikan data dengan tepat secara konsisten. Rata-rata

makro (*macro average*) untuk *precision*, *recall*, dan *f1-score* juga berada di angka 0,97, yang menunjukkan bahwa performa model stabil pada semua kelas tanpa adanya bias terhadap kelas tertentu. Rata-rata berbobot (*weighted average*) yang sama dengan nilai makro mencerminkan bahwa distribusi data yang seimbang tidak mempengaruhi performa model. Secara keseluruhan, kombinasi arsitektur CNN (Conv1D) dan BiLSTM memberikan hasil klasifikasi yang optimal, memanfaatkan kekuatan CNN (Conv1D) dalam ekstraksi fitur spasial dan BiLSTM dalam menangkap urutan temporal untuk mengidentifikasi pola kompleks dalam data sekuensial.

	precision	recall	f1-score	support
False	0.96	0.98	0.97	8367
True	0.98	0.95	0.97	8367
accuracy			0.97	16734
macro avg	0.97	0.97	0.97	16734
weighted avg	0.97	0.97	0.97	16734

Gambar 10. *Classification Report*

IV. KESIMPULAN

Hasil penerapan SMOTE mampu menyeimbangkan data yang semula tidak seimbang, di mana jumlah transaksi pembelian (kelas "True") jauh lebih sedikit dibandingkan dengan transaksi non-pembelian (kelas "False"). Sebelum diterapkannya SMOTE, jumlah sampel kelas "True" hanya sekitar 1.497 data dibandingkan lebih dari 8.000 data untuk kelas "False". Ketidakseimbangan ini berpotensi membuat model pembelajaran mesin bias terhadap kelas mayoritas. Dengan menerapkan SMOTE, distribusi data menjadi seimbang, sehingga model memiliki kesempatan yang sama dalam mempelajari karakteristik kedua kelas. Setelah proses balancing data, model CNN (Conv1D) + BiLSTM yang diterapkan menunjukkan performa yang sangat baik, dengan akurasi mencapai 97% serta nilai precision, recall, dan f1-score yang tinggi dan konsisten untuk kedua kelas. Grafik akurasi dan loss selama pelatihan menunjukkan peningkatan performa yang stabil tanpa tanda-tanda overfitting yang signifikan. Hasil ini menunjukkan bahwa model dapat mengenali pola transaksi dengan cukup baik. Metode SMOTE dapat meningkatkan kompleksitas komputasi karena menghasilkan data sintetis dalam jumlah besar. Selain itu, hasil penelitian

bergantung pada pengaturan hyperparameter model yang belum sepenuhnya dioptimalkan. Penelitian selanjutnya disarankan untuk melakukan tuning hyperparameter agar performa model lebih optimal serta menguji model pada dataset baru dengan fitur yang berbeda untuk meningkatkan generalisasi model. Secara keseluruhan, penerapan metode SMOTE dan arsitektur CNN (Conv1D) + LSTM telah berhasil meningkatkan akurasi prediksi transaksi pembelian online dan mengurangi bias data.

DAFTAR PUSTAKA

- [1]. Abdullah-All-Tanvir, Ali Khandokar, I., Muzahidul Islam, A. K. M., Islam, S., & Shatabda, S. (2023). A gradient boosting classifier for purchase intention prediction of online shoppers. *Heliyon*, 9(4). <https://doi.org/10.1016/j.heliyon.2023.e15163>
- [2]. Angraini, N. A., & Lhaksmana, K. M. (2024). Sentiment Analysis About Legislative Elections using Deep Learning with LSTM and CNN Models. *Technology and Science (BITS)*, 6(1). <https://doi.org/10.47065/bits.v6i1.5283>
- [3]. Baati, K., & Mohsil, M. (2020). Real-time prediction of online shoppers' purchasing intention using random forest. *IFIP Advances in Information and Communication Technology*, 583 IFIP, 43–51. https://doi.org/10.1007/978-3-030-49161-1_4
- [4]. Chetana R, A Shubha Rao, & Mahantesh K. (2023). Application of Conv-1D and Bi-LSTM to Classify. *(IJACSA) International Journal of Advanced Computer Science and Applications*, Vol. 14(No. 6).
- [5]. Dharmesti, M., Dharmesti, T. R. S., Kuhne, S., & Thaichon, P. (2019). Understanding online shopping behaviours and purchase intentions amongst millennials. *Young Consumers*, 22(1), 152–167. <https://doi.org/10.1108/YC-12-2018-0922>
- [6]. Du, L., Gao, F., Chen, X., Jia, R., Wang, J., Zhang, J., Han, S., & Zhang, D. (2021). TabularNet: A Neural Network Architecture for Understanding Semantic Structures of Tabular Data. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 322–331. <https://doi.org/10.1145/3447548.3467228>

- [7]. Grimaldo, J. R., Uy, C., & Manalo, R. A. (2023). Structural Equation Model On Determinants Of Intention To Use E-Commerce. *Journal of Computational Innovation and Analytics (JCIA)*, 2(2), 145–174. <https://doi.org/10.32890/jcia2023.2.2.2>
- [8]. Kulkarni, A. D. (2024). Fuzzy Convolution Neural Networks for Tabular Data Classification. *IEEE Access*, 12, 151846–151855. <https://doi.org/10.1109/ACCESS.2024.3479882>
- [9]. Le-Hoang, P. V. (2020). Factors affecting online purchase intention: the case of e-commerce on lazada. *Independent Journal of Management & Production*, 11(3), 1018–1033. <https://doi.org/10.14807/ijmp.v11i3.1088>
- [10]. Lin, B., & Shen, B. (2023). Study of Consumers' Purchase Intentions on Community E-commerce Platform with the SOR Model: A Case Study of China's "Xiaohongshu" App. *Behavioral Sciences*, 13(2). <https://doi.org/10.3390/bs13020103>
- [11]. Mamani, E. L. (2023). Expanded Convolutional Network for Tabular Data. Submission and Formatting Instruction for ICML.
- [12]. Muhammad, H. (2022). Online Shopper Intention Analysis Using Conventional Machine Learning and Deep Neural Network Classification Algorithm. Membandingkan Algoritma Machine Learning Konvensional Seperti Decision Tree, Logistic Regression, Dan Random Forest Dengan Deep Neural Network (DNN).
- [13]. Nugroho H, Y. D., Zakiyabarsi, F., & Paramita, A. J. (2025). Implementasi SMOTE-Enn Dan Borderline Smote Terhadap Performa LightGBM Pada Imbalanced Class. *Rabit: Jurnal Teknologi Dan Sistem Informasi Univrab*, 10(1), 51–59. <https://doi.org/10.36341/rabit.v10i1.5436>
- [14]. Purnama, A., Maulana Yusup, A., Wibowo, A., & Susilawati, D. (2021). Uji Algoritma Random Forest Pada Dataset Online Shoppers Purchasing Intention. *Jurnal IKRA-ITH Informatika*, 5(1).
- [15]. Torres, D., Mx, A., & Cepeda, L. K. (2024). Machine Learning for Predicting Online Shoppers' Purchase Intentions. <https://doi.org/10.13140/RG.2.2.15297.36968>
- [16]. Wang, R. (2023). Research And Analysis of Online Shopper Intention. In *Journal of Education, Humanities and Social Sciences EAIS* (Vol. 2023).
- [17]. Wen, Z., Lin, W., & Liu, H. (2023). Machine-Learning-Based Approach for Anonymous Online Customer Purchase Intentions Using Clickstream Data. *Systems*, 11(5). <https://doi.org/10.3390/systems11050255>
- [18]. Wijayanti, N. P. Y. T. (2021). SMOTE: POTENSI DAN KEKURANGANNYA PADA SURVEI. *E-Jurnal Matematika*, 10(4), 235. <https://doi.org/10.24843/mtk.2021.v10.i04.p348>
- [19]. Yadav, S., & Student, M. T. (2023). Prediction Of Online Shopper's Buying Intention Using Algorithms Of Pyspark Mllib. In *International Journal of Current Science (IJCS PUB)* www.ijcs.pub.org (Vol. 13, Issue 2). www.ijcs.pub.org
- [20]. Yoke Cheng, S., Alisa Hussain, I., Apparavu, K., & Trianna Rosli, N. (2019). Factors Influencing Online Shopping Intention Among Malaysians: A Quantitative-Based Study. *Electronic Journal of Business and Management*, 7, 2550–1380.
- [21]. Yuliana, R., Hamdani, M., Tinggi, S., Bank, I. E., & Jateng, B. (2023). STUDY OF ONLINE BUYING INTENTION STUDY OF ONLINE SHOPPING. *Business and Accounting Research (IJEBA)* *Peer Reviewed-International Journal*, 13(103). <https://jurnal.stie-aas.ac.id/index.php/IJEBA>