

APPLICATION OF MACHINE LEARNING AND DEEP LEARNING TO PREDICT FINANCIAL PRODUCT SUBSCRIPTIONS BASED ON CUSTOMER FEATURES

PENERAPAN MACHINE LEARNING DAN DEEP LEARNING UNTUK MEMPREDIKSI LANGGANAN PRODUK KEUANGAN BERDASARKAN FITUR PELANGGAN

Harditya Estu Prayoga¹, De Rosal Ignatus Moses Setiadi², Hendy Kurniawan³

Universitas Dian Nuswantoro, Jl. Imam Bonjol No.207, Semarang, Indonesia

111202113683@mhs.dinus.ac.id¹, moses@dsn.dinus.ac.id², 111202113620@mhs.dinus.ac.id³

Abstract - The financial industry faces challenges in predicting consumer behaviour, especially in forecasting decisions related to subscribing to financial products like term deposits. This study applies machine learning and deep learning to predict subscriptions based on demographic and behavioural data from the Bank Marketing dataset from the UCI Machine Learning Repository. The models tested include Support Vector Machine (SVM), Logistic Regression (LR), Random Forest (RF), Bidirectional Gated Recurrent Unit (BiGRU), and Bidirectional Long Short-Term Memory (BiLSTM). Model performance is evaluated using metrics such as accuracy, precision, recall, F1-score, and confusion matrix. Results show that BiGRU achieves the highest accuracy of 92.52%, outperforming other models, with SVM and BiLSTM also showing strong performance. However, all models still face limitations in detecting subscribing customers, as evidenced by the high false negative rate. These findings highlight the potential of machine learning and deep learning to support data-driven decision-making in financial marketing, despite limitations such as the use of a single data source and the lack of consideration for external factors affecting customer decisions.

Keywords - Financial Industry, Machine Learning, Deep Learning, Subscription Prediction, Bank Marketing dataset.

Abstrak - Industri keuangan menghadapi tantangan dalam memprediksi perilaku konsumen, khususnya dalam meramalkan keputusan berlangganan produk keuangan seperti deposito berjangka. Penelitian ini menerapkan machine learning dan deep learning untuk memprediksi langganan berdasarkan data demografis dan perilaku nasabah menggunakan dataset Bank Marketing dari UCI Machine Learning Repository. Model yang diuji meliputi Support Vector Machine (SVM), Logistic Regression (LR), Random Forest (RF), Bidirectional Gated Recurrent Unit (BiGRU), dan Bidirectional Long Short-Term Memory (BiLSTM). Evaluasi kinerja dilakukan menggunakan metrik akurasi, presisi, recall, F1-score, dan confusion matrix. Hasil menunjukkan BiGRU memiliki akurasi terbaik sebesar 92,52%, mengungguli model lainnya, dengan SVM dan BiLSTM juga menunjukkan performa yang baik. Namun, semua model masih mengalami keterbatasan dalam mendeteksi pelanggan yang berlangganan, tercermin dari tingginya false negative. Temuan ini menunjukkan potensi machine learning dan deep learning dalam mendukung keputusan pemasaran berbasis data di sektor perbankan, meskipun dengan keterbatasan seperti penggunaan satu sumber data dan tidak mempertimbangkan faktor eksternal yang memengaruhi keputusan pelanggan.

Kata Kunci - Industri Keuangan, Machine Learning, Deep Learning, Prediksi Langganan, Dataset Bank Marketing.

I. PENDAHULUAN

Sektor keuangan menghadapi tantangan besar dalam memahami perilaku pelanggan dan memprediksi keputusan mereka terkait langganan produk keuangan seperti deposito berjangka. Meskipun produk deposito berjangka menawarkan tingkat pengembalian yang stabil, tingkat adopsinya sangat bergantung pada berbagai faktor, termasuk karakteristik demografis dan perilaku finansial nasabah. Oleh karena itu, diperlukan sistem prediksi yang mampu mengidentifikasi nasabah yang memiliki potensi tinggi untuk berlangganan produk tersebut. Pertanyaan utama dalam penelitian ini adalah bagaimana penerapan metode machine learning dan deep learning dapat meningkatkan akurasi prediksi langganan produk keuangan berdasarkan data pelanggan? Dengan menjawab pertanyaan ini, penelitian ini bertujuan untuk mengembangkan sistem prediksi yang lebih efektif dalam menentukan pelanggan yang berpotensi berlangganan produk deposito berjangka.

Penggunaan machine learning dan deep learning telah terbukti memberikan hasil yang lebih baik dibandingkan dengan teknik konvensional dalam memproses volume data besar dengan akurasi yang lebih tinggi. Berbagai studi sebelumnya telah meneliti penggunaan machine learning dalam memprediksi langganan produk keuangan. Sebagai contoh, penelitian oleh Riyyasy et al. (2023) memanfaatkan algoritma seperti Random Forest dan Support Vector Machine yang digunakan untuk memprediksi langganan produk term deposit berdasarkan fitur-fitur demografis dan riwayat transaksi nasabah [1]. Selain itu, penelitian oleh Syafa'ah dan Lestandy (2021) mengenai prediksi kasus aktif Covid-19 dengan deep learning menunjukkan bagaimana model ini efektif dalam menangani data yang kompleks dan dinamis[2].

Dampak ekonomi dan operasional dari prediksi langganan yang akurat sangat signifikan bagi industri keuangan. Dengan memprediksi pelanggan yang berpotensi berlangganan, lembaga keuangan dapat mengoptimalkan upaya pemasaran mereka dan meningkatkan tingkat konversi. Hal ini dapat berujung pada peningkatan profitabilitas dan efisiensi operasional, karena sumber daya dapat difokuskan pada pelanggan yang lebih mungkin untuk berlangganan produk keuangan tersebut.

Penggunaan machine learning dan deep learning juga meluas ke bidang lain seperti deteksi kecurangan dan prediksi bencana alam. Tarissa dan Dewayanto (2024) mengungkapkan bahwa teknik ini dapat meningkatkan akurasi dalam mendeteksi fraud kartu kredit[3]. Sementara itu, Prasetyo dan Dewayanto (2024) menyoroti pemanfaatan teknologi ini dalam mendeteksi kecurangan laporan keuangan melalui pendekatan machine learning, deep learning, dan data mining [4]. Pitaloka et al. (2024) pun menerapkan machine learning dalam memprediksi bencana banjir, menggarisbawahi pentingnya analisis prediktif berbasis big data[5].

Dalam penelitian ini, kami berfokus pada penerapan machine learning dan deep learning untuk memprediksi langganan produk deposito berjangka menggunakan fitur-fitur nasabah yang tersedia. Diharapkan, hasil dari penelitian ini dapat memberikan kontribusi yang berarti dalam pengembangan sistem prediksi yang lebih akurat di sektor perbankan, serta membantu lembaga keuangan dalam menyusun strategi pemasaran yang lebih terarah dan berbasis data.

II. SIGNIFIKASI STUDI

A. Studi Literatur

Beberapa penelitian sebelumnya telah mengkaji penerapan machine learning dan deep learning dalam analisis data keuangan, yang relevansinya dapat diterapkan pada prediksi langganan produk keuangan. Riyyasy et al. (2023) menggunakan Random Forest dan Support Vector Machine untuk memprediksi langganan produk term deposit, menekankan pentingnya fitur pelanggan seperti data demografis dan riwayat transaksi. Namun, akurasi prediksi masih terbatas dalam mengidentifikasi nasabah yang benar-benar berlangganan [1].

Syafa'ah dan Lestandy (2021) menerapkan deep learning untuk memprediksi kasus Covid-19, yang menunjukkan kemampuan menangani data dinamis. Meskipun tidak langsung berkaitan dengan produk keuangan, metode ini dapat diadaptasi untuk menganalisis pola perilaku nasabah, meskipun studi terdahulu belum sepenuhnya menggali potensi deep learning dalam sektor keuangan [2].

Tarissa dan Dewayanto (2024) menggunakan machine learning dan deep learning untuk deteksi penipuan kartu kredit, relevan untuk prediksi langganan produk keuangan. Namun, penelitian sebelumnya lebih fokus pada data transaksi dan belum mengeksplorasi data perilaku nasabah secara luas, yang menjadi gap yang ingin diisi penelitian ini [3]. Prasetyo dan Dewayanto (2024) menyoroti penggunaan teknik-teknik ini untuk mendeteksi kecurangan laporan keuangan, namun belum mengkaji prediksi perilaku konsumen dalam memilih produk keuangan. Penelitian ini bertujuan mengisi celah tersebut dengan fokus pada prediksi langganan produk keuangan [4].

Pitaloka et al. (2024) menggunakan machine learning untuk memprediksi bencana banjir, meskipun bukan dalam konteks keuangan. Penelitian ini mengilustrasikan penerapan prediksi, namun penerapan serupa dalam konteks keuangan untuk memprediksi langganan produk keuangan masih terbatas [5]. Luthfiansyah et al. (2023) menggabungkan deep learning dan regresi linier untuk memprediksi pergerakan saham, yang relevan dengan analisis perilaku nasabah, namun penerapan gabungan metode ini dalam prediksi langganan produk keuangan masih jarang [6].

Penelitian ini memberikan kontribusi signifikan dengan menggunakan model deep learning canggih seperti BiGRU dan BiLSTM untuk memprediksi langganan produk keuangan, mengisi celah dalam literatur. Temuan ini dapat mendorong penelitian lebih lanjut di sektor keuangan dan memberikan wawasan baru untuk pengembangan strategi pemasaran berbasis data.

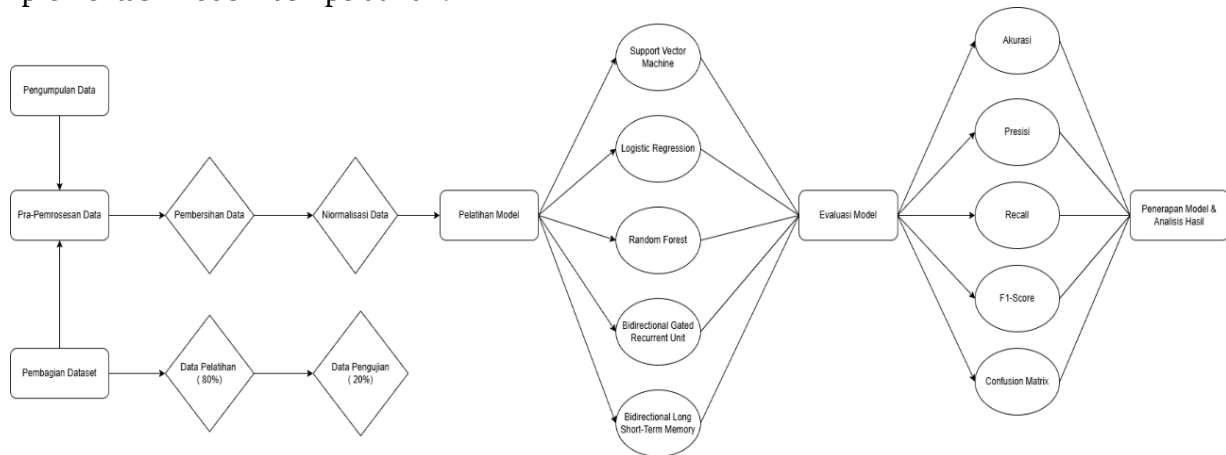
B. Data Penelitian

Penelitian ini menggunakan dataset Bank Marketing dari UCI Machine Learning Repository, yang mencakup 41.188 entri dengan 21 kolom, berisi informasi tentang kampanye pemasaran langsung bank di Portugal melalui telepon untuk mempromosikan produk deposito berjangka. Dataset ini mencakup karakteristik nasabah seperti usia, pekerjaan, pendidikan, status pernikahan, serta riwayat interaksi dengan kampanye sebelumnya. Variabel target adalah *y*, yang menunjukkan apakah nasabah berlangganan produk atau tidak. Selain itu, terdapat fitur seperti jumlah kontak, durasi panggilan, dan riwayat keterlibatan sebelumnya. Pra-pemrosesan dilakukan dengan menangani nilai hilang dan mengubah variabel kategorikal menjadi numerik menggunakan one-hot encoding. Dataset ini digunakan untuk membangun model prediktif guna memahami perilaku nasabah dalam memilih produk keuangan, sebagaimana juga dijelaskan oleh Religia et al. (2021) [7].

C. Metode Penelitian

1. Metodologi

Penelitian ini disusun dengan tujuan untuk memprediksi langganan produk deposito berjangka melalui penerapan berbagai algoritma machine learning dan deep learning [8]. Proses penelitian ini mencakup beberapa tahapan penting, yaitu pengumpulan data, pra-pemrosesan data, pembagian dataset menjadi data latih dan data uji, pelatihan model, evaluasi performa model, serta implementasi model hasil pelatihan.



Gambar 1. Alur Metode Penelitian

Metodologi dalam penelitian ini mencakup serangkaian langkah utama dalam menerapkan machine learning dan deep learning untuk memprediksi keputusan langganan produk keuangan [9]. Proses dimulai dengan pengambilan data dari dataset Bank Marketing yang tersedia di UCI Machine Learning Repository. Data tersebut kemudian melalui tahap pra-pemrosesan, yang meliputi pembersihan untuk menangani nilai yang hilang, penghapusan data duplikat, serta normalisasi agar semua fitur memiliki skala yang seragam. Setelah itu, data dibagi menjadi 80% untuk pelatihan dan 20% untuk pengujian. Pada tahap pelatihan, beberapa algoritma diterapkan, seperti Support Vector Machine (SVM), Logistic Regression (LR), Random Forest (RF), Bidirectional Gated Recurrent Unit (BiGRU), dan Bidirectional Long Short-Term Memory (BiLSTM). Setelah proses pelatihan selesai, model dievaluasi menggunakan berbagai metrik, seperti akurasi, precision, recall, F1-score, dan confusion matrix, untuk menilai kinerjanya. Hasil evaluasi dari masing-masing model dibandingkan untuk memilih algoritma yang paling efektif, yang kemudian digunakan untuk memprediksi dengan lebih tepat apakah seorang nasabah akan berlangganan produk deposito berjangka.

2. Ilustrasi Dataset

Untuk memperjelas pemahaman tentang dataset yang digunakan, berikut disajikan visualisasi struktur dari dataset tersebut:

1	age	job	marital	education	default	housing	loan	contact	month	day_of_w	duration	campaign	pdays	previous	poutcome	emp.var.rz	cons.price	cons.conf	euribor3m	nr.employ	y
2	56	housemaid	married	basic.4y	no	no	no	telephone	may	mon	261	1	999	0	nonexister	1.1	93.994	-36.4	4.857	5191	no
3	57	services	married	high.schoc	unknown	no	no	telephone	may	mon	149	1	999	0	nonexister	1.1	93.994	-36.4	4.857	5191	no
4	37	services	married	high.schoc	no	yes	no	telephone	may	mon	226	1	999	0	nonexister	1.1	93.994	-36.4	4.857	5191	no
5	40	admin.	married	basic.6y	no	no	no	telephone	may	mon	151	1	999	0	nonexister	1.1	93.994	-36.4	4.857	5191	no

Gambar 2. Bank additional.csv

Dataset ini memuat informasi terkait nasabah, meliputi usia, jenis pekerjaan, status pernikahan, tingkat pendidikan, serta kepemilikan pinjaman. Selain itu, dataset ini juga mencakup data mengenai kampanye pemasaran sebelumnya, durasi interaksi, dan kondisi ekonomi saat kampanye berlangsung. Variabel target dalam dataset ini adalah y, yang merepresentasikan apakah seorang nasabah memutuskan untuk berlangganan produk deposito berjangka. Data tersebut dimanfaatkan

untuk menganalisis berbagai faktor yang berpengaruh terhadap keputusan nasabah dalam memilih produk keuangan.

3. Proses awal Data

Tahap proses awal data dimulai dengan penanganan data hilang dan duplikasi, di mana nilai yang hilang diisi menggunakan teknik imputasi, seperti mengganti dengan nilai median atau modus [10]. Variabel kategorikal, seperti pekerjaan, status pernikahan, pendidikan, perumahan, pinjaman, kontak, dan hasil pemasaran, diubah menjadi format numerik melalui one-hot encoding atau label encoding. Selanjutnya, data numerik seperti usia, durasi, jumlah kampanye, dan jumlah kontak sebelumnya dinormalisasi untuk memastikan semua fitur memiliki skala yang seragam. Akhirnya, dataset dibagi menjadi 80% untuk pelatihan dan 20% untuk pengujian agar model dapat dilatih dan diuji menggunakan data yang terpisah.

4. Metode Prediksi Langganan Produk Keuangan

Dataset dibagi menjadi dua bagian: *training set* dan *testing set*. Data latih (*training set*) digunakan untuk melatih model, sementara data uji (*testing set*) digunakan untuk *mengevaluasi* kinerja model. Pembagian ini dilakukan dengan *rasio* 80:20, di mana 80% data digunakan untuk pelatihan dan 20% untuk pengujian model.

a. Support Vector Machine (SVM)

SVM memiliki rumus matematis terkait dengan pencarian hyperplane yang memisahkan kelas-kelas dalam ruang fitur [11]. Konsep dasar dari SVM adalah mencari hyperplane optimal yang memaksimalkan margin antara dua kelas. Rumus dasar dari SVM adalah optimisasi masalah convex yang dapat dijelaskan sebagai:

$$\min \frac{1}{2} \|w\|^2 \quad \text{s.t.} \quad y_i(w^T x_i + b) \geq 1, \forall i$$

Gambar 3. Rumus SVM

b. Logistic Regression (LR)

Logistic Regression digunakan untuk memprediksi probabilitas, dan model ini menghasilkan fungsi logistik (sigmoid) untuk memetakan nilai kontinu ke probabilitas [12]. Rumus fungsi logistik adalah:

$$p(y = 1|x) = \frac{1}{1 + e^{-(w^T x + b)}}$$

Gambar 4. Rumus LR

c. Random Forest (RF)

Random Forest merupakan pendekatan ensemble learning yang berasal dari teknik CART (Classification and Regression Trees). Metode ini berfungsi dengan menggabungkan beberapa pohon keputusan (Decision Tree) untuk meningkatkan akurasi prediksi dan mengurangi risiko overfitting, di mana keputusan akhir ditentukan oleh suara mayoritas dari seluruh pohon yang ada [13]. Model ini membangun N pohon keputusan secara acak, kemudian menghasilkan prediksi berdasarkan pemungutan suara mayoritas dari pohon-pohon tersebut:

$$\hat{y} = \text{mode}(h_i(x)_{i=1}^N)$$

Gambar 5. Rumus RF

d. Bidirectional Gated Recurrent Unit (BIGRU)

BIGRU adalah model RNN yang lebih kompleks, dan meskipun tidak ada rumus tunggal yang menggambarkan keseluruhan prosesnya, prinsip kerjanya didasarkan pada fungsi pembaruan gate untuk memproses urutan data [14]. Formulasnya adalah gabungan dari beberapa langkah-langkah dalam jaringan RNN, tetapi untuk BIGRU, pembaruan dilakukan di dua arah (maju dan mundur). Rumus dasar untuk Gated Recurrent Unit (GRU) adalah:

$$\begin{aligned} r_t &= \sigma(W_r x_t + U_r h_{t-1}) \\ z_t &= \sigma(W_z x_t + U_z h_{t-1}) \\ \tilde{h}_t &= \tanh(W_h x_t + U_h(r_t \circ h_{t-1})) \\ h_t &= (1 - z_t) \circ h_{t-1} + z_t \circ \tilde{h}_t \end{aligned}$$

Gambar 6. Rumus BiGRU

e. Bidirectional Long Short-Term Memory (BiLSTM)

BiLSTM adalah varian dari LSTM (Long Short-Term Memory), dan tidak ada rumus tunggal untuk keseluruhan model [15]. LSTM itu sendiri memiliki beberapa komponen rumus seperti input gate, forget gate, dan output gate. Untuk BiLSTM, dua LSTM digunakan: satu berjalan dari awal ke akhir (maju), dan satu lagi berjalan dari akhir ke awal (mundur). Rumus dasar dari LSTM adalah:

$$\begin{aligned} f_t &= \sigma(W_f x_t + U_f h_{t-1} + b_f) \\ i_t &= \sigma(W_i x_t + U_i h_{t-1} + b_i) \\ \tilde{C}_t &= \tanh(W_C x_t + U_C h_{t-1} + b_C) \\ C_t &= f_t \circ C_{t-1} + i_t \circ \tilde{C}_t \\ o_t &= \sigma(W_o x_t + U_o h_{t-1} + b_o) \\ h_t &= o_t \circ \tanh(C_t) \end{aligned}$$

Gambar 7. Rumus BiLSTM

f. Evaluasi

Model pembelajaran mesin ini dievaluasi dengan menggunakan berbagai metrik kinerja untuk mengukur efektivitas dalam memprediksi Langganan Produk Keuangan, yaitu:

$$\begin{aligned} \text{Akurasi} &= \frac{\text{True Positives} + \text{True Negatives}}{\text{True Positives} + \text{True Negatives} + \text{False Positives} + \text{False Negatives}} \\ \text{Presisi} &= \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \\ \text{Recall} &= \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \\ \text{F1-score} &= 2 \times \frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} \end{aligned}$$

Gambar 8. Metrik Performa

Evaluasi model dilakukan untuk menilai kinerja algoritma dalam memprediksi langganan produk deposito berjangka. Berbagai metrik digunakan dalam proses ini, termasuk akurasi, yang menunjukkan persentase prediksi yang benar; presisi, yang mengukur proporsi prediksi positif yang akurat; serta recall, yang menilai kemampuan model dalam menangkap seluruh instance positif [16]. F1-score digunakan untuk menyeimbangkan antara presisi dan recall, terutama pada dataset yang tidak seimbang. Selain itu, confusion matrix memberikan gambaran lebih rinci tentang distribusi hasil prediksi, mencakup jumlah True Positive, True Negative, False Positive, dan False Negative. Berdasarkan hasil evaluasi dari semua metrik tersebut, model dengan kinerja terbaik dipilih untuk menghasilkan prediksi yang lebih akurat.

III. HASIL DAN PEMBAHASAN

Secara keseluruhan, model BiGRU menampilkan hasil terbaik dalam hal akurasi, meskipun diikuti oleh SVM dan Logistic Regression (LR), yang juga memberikan performa tinggi. Model deep learning seperti BiGRU dan BiLSTM memberikan keuntungan lebih dalam menangani data berurutan, yang menjadikannya pilihan yang sangat baik untuk prediksi langganan produk keuangan. Keunggulan BiGRU dan BiLSTM dapat dijelaskan oleh kemampuan kedua model ini dalam menangani data dengan urutan waktu yang kompleks dan pola-pola temporal yang sulit ditangkap oleh model-model tradisional seperti SVM dan LR. Selain itu, kemampuan BiGRU dan BiLSTM untuk menangani dependensi jangka panjang antar fitur dalam data transaksi nasabah memungkinkan model ini untuk menangkap pola yang lebih kompleks yang berkontribusi pada akurasi yang lebih tinggi.

A. Hasil Model Evaluasi

TABEL I. DECISION TREE

	Precision	Recal	F1-Score	Support
0	0.92	0.98	0.95	7303
1	0.68	0.30	0.42	935
Accuracy			0.90	8238
Macro AVG	0.80	0.64	0.68	8238
Weighted AVG	0.89	0.90	0.89	8238

SVM Accuracy 90,46%

TABEL II. LOGISTIC REGRESSION

	Precision	Recal	F1-Score	Support
0	0.92	0.97	0.94	7303
1	0.60	0.30	0.40	935
Accuracy			0.90	8238
Macro AVG	0.76	0.64	0.67	8238
Weighted AVG	0.88	0.90	0.88	8238

Logistic Regression Accuracy 89,79%

TABEL III. RANDOM FOREST

	Precision	Recal	F1-Score	Support
0	0.94	0.94	0.94	7303
1	0.53	0.50	0.51	935
Accuracy			0.89	8238
Macro AVG	0.73	0.72	0.72	8238
Weighted AVG	0.89	0.89	0.89	8238

Random Forest Accuracy 89,22%

TABEL IV. BiGRU

	Precision	Recal	F1-Score	Support
0	0.93	0.96	0.94	7303
1	0.58	0.47	0.52	935
Accuracy			0.90	8238
Macro AVG	0.75	0.71	0.73	8238
Weighted AVG	0.89	0.90	0.90	8238

BiGRU Accuracy 92,52%

TABEL V. BiLSTM

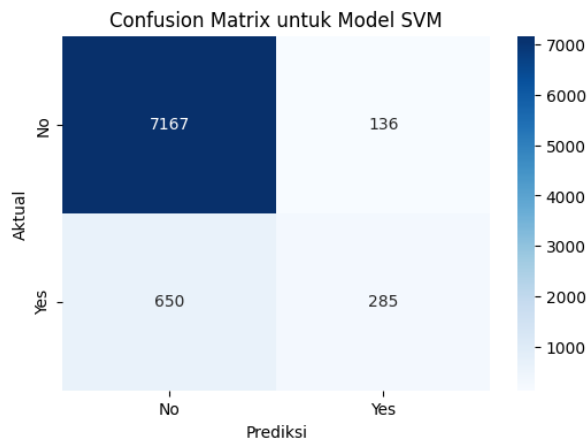
	Precision	Recal	F1-Score	Support
0	0.92	0.98	0.95	7303
1	0.63	0.32	0.43	935
Accuracy			0.90	8238
Macro AVG	0.77	0.65	0.69	8238
Weighted AVG	0.89	0.90	0.89	8238

BiGRU Accuracy 92,24%

B. Confusion Matrix

Analisis confusion matrix memberikan wawasan mengenai distribusi prediksi model terhadap kelas yang sesungguhnya. Confusion matrix mengukur empat metrik utama: True Positives (TP), True Negatives (TN), False Positives (FP), dan False Negatives (FN), yang digunakan untuk mengevaluasi sejauh mana model berhasil memprediksi keputusan langganan produk. Setiap model yang diuji menunjukkan kelebihan dan kekurangan dalam mengidentifikasi nasabah yang berpotensi berlangganan produk keuangan, yang tercermin pada hasil confusion matrix masing-masing model. Misalnya, meskipun BiGRU dan BiLSTM menunjukkan akurasi yang lebih tinggi secara keseluruhan, mereka juga cenderung menghasilkan False Positives lebih banyak dibandingkan dengan model lain, seperti SVM, yang lebih berhati-hati dalam memprediksi nasabah berlangganan.

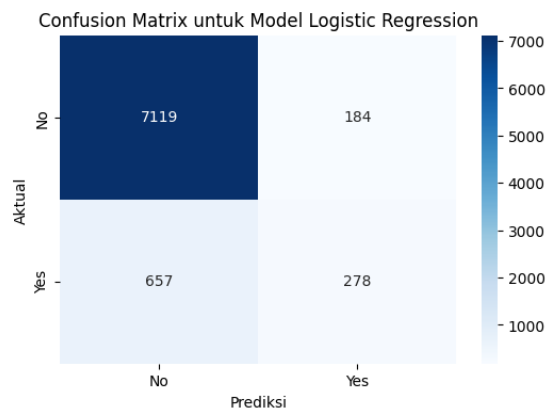
1. SVM



Gambar 9 Confusion Matrix SVM

Pada algoritma SVM (akurasi 90.46%) memiliki performa yang baik dalam mengidentifikasi nasabah yang tidak berlangganan (TN tinggi), namun kesulitan mendeteksi nasabah yang berlangganan (FN tinggi).

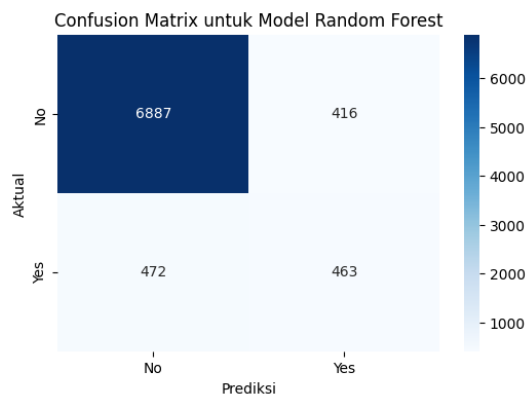
2. Logistic Regression



Gambar 10 Confusion Matrix LR

Logistic Regression (akurasi 89.79%) juga efektif mengidentifikasi nasabah yang tidak berlangganan, tetapi memiliki FN yang tinggi, menunjukkan kesulitan dalam mendeteksi nasabah yang berpotensi berlangganan.

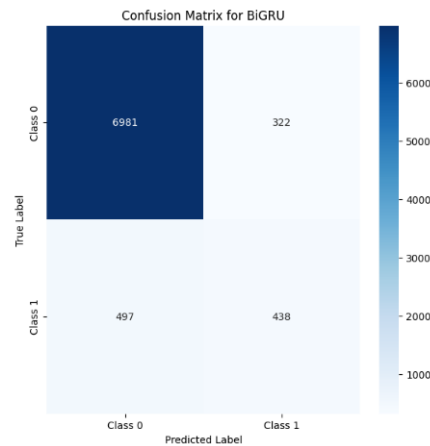
3. Random Forest



Gambar 11 Confusion Matrix RF

Random Forest (akurasi 89.22%) cukup seimbang antara TN dan TP, namun memiliki kesalahan lebih banyak dalam mendeteksi nasabah yang berlangganan (FN lebih tinggi).

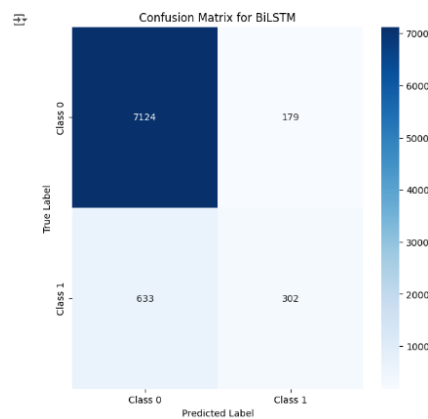
4. BiGRU



Gambar 12 Confusion Matrix BiGRU

BiGRU (akurasi 92.52%) unggul dalam mendeteksi nasabah yang berlangganan (TP lebih tinggi), meskipun masih terdapat beberapa FN.

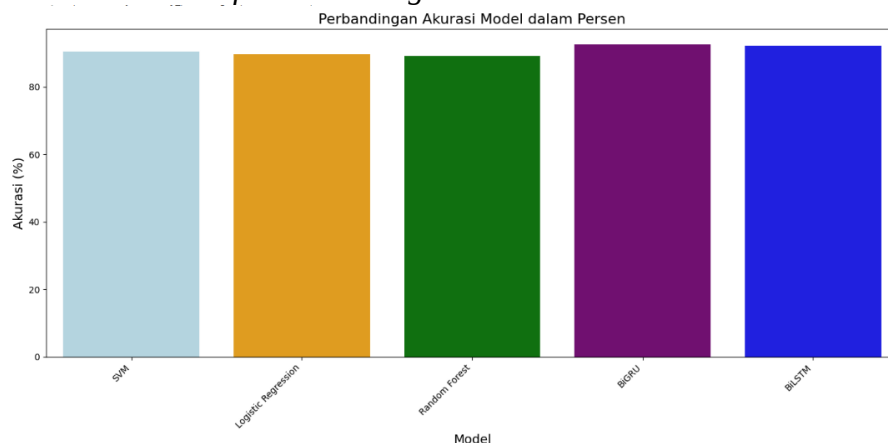
5. BiLSTM



Gambar 13 Confusion Matrix BiLSTM

BiLSTM (akurasi 92.24%) efektif memprediksi nasabah yang tidak berlangganan, namun memiliki FN yang cukup tinggi, menunjukkan kesulitan mendeteksi nasabah yang berpotensi berlangganan.

C. Visualisasi Model dan Grafik Perbandingan



Gambar 14 Grafik dan Visualisasi Model

Pada grafik Perbandingan Akurasi Model, BiGRU dan BiLSTM memiliki akurasi tertinggi sekitar 92%, diikuti oleh SVM (90.46%) dan Logistic Regression (89.79%), dengan Random Forest sedikit lebih rendah (89.22%). Meskipun BiGRU dan BiLSTM unggul dalam akurasi, SVM menunjukkan performa lebih baik dalam Precision, yaitu kemampuan mengidentifikasi nasabah yang berlangganan dengan lebih sedikit prediksi palsu positif. Grafik ini juga membandingkan model dalam metrik lain seperti Precision, Recall, dan F1 Score. BiGRU dan BiLSTM menonjol dalam Recall, yang berarti mereka lebih efektif dalam menangkap lebih banyak nasabah yang berpotensi berlangganan. Ini penting dalam prediksi langganan produk keuangan karena meningkatkan peluang konversi. Namun, meskipun BiGRU dan BiLSTM unggul dalam Recall dan akurasi, SVM menunjukkan performa baik dalam Precision. Perbandingan ini menunjukkan bahwa tidak ada model yang dominan di semua metrik. BiGRU dan BiLSTM unggul dalam Recall dan akurasi, sementara SVM lebih unggul dalam Precision. Pemilihan model harus mempertimbangkan trade-off antara Recall dan Precision, tergantung pada prioritas aplikasi. Jika fokus pada mengidentifikasi lebih banyak nasabah berpotensi berlangganan (tinggi Recall), maka BiGRU dan BiLSTM lebih baik, namun jika prioritasnya adalah meminimalkan prediksi palsu positif (tinggi Precision), SVM lebih tepat. Perbandingan ini juga ditempatkan dalam konteks penelitian sebelumnya, seperti Riyyasy et al. (2023), yang menunjukkan bahwa model deep learning seperti BiGRU dan BiLSTM lebih akurat dibandingkan model konvensional seperti Random Forest dan SVM, menegaskan kekuatan deep learning dalam menangani data besar dan kompleks di sektor keuangan.

IV. KESIMPULAN

Penelitian ini berhasil menerapkan machine learning dan deep learning untuk memprediksi langganan produk deposito berjangka, dengan model BiGRU dan BiLSTM menunjukkan akurasi lebih dari 92%, mengungguli model lain seperti SVM, Logistic Regression, dan Random Forest. Meskipun demikian, model ini masih memiliki kelemahan dalam mendeteksi nasabah berpotensi berlangganan, tercermin dari tingginya False Negatives. BiGRU tampil optimal dalam akurasi, recall, dan F1-score, namun perlu pengurangan kesalahan False Negatives melalui modifikasi model atau teknik oversampling. Temuan ini menunjukkan efektivitas machine learning dan deep learning dalam memprediksi keputusan nasabah, dengan potensi besar untuk aplikasi dalam strategi pemasaran keuangan berbasis data. Penelitian selanjutnya disarankan untuk mengeksplorasi teknik ensemble atau data dinamis, serta penerapan dalam sektor keuangan lainnya. Hasil penelitian ini dapat digunakan oleh praktisi perbankan untuk merancang strategi pemasaran yang lebih akurat dan meningkatkan retensi pelanggan.

REFERENSI

- [1] L. S. and M. L. , “Penerapan Deep Learning untuk Prediksi Kasus Aktif Covid-19,” *Jurnal Sains Komputer & Informatika*, vol. 5, pp. 453-457, 2021.
- [2] M. . R. A. Riyyasy, W. N. Aghniya and H. T. , “Penerapan Algoritma Machine Learning Untuk Memprediksi Term Deposit Nasabah Perbankan,” *LEDGER: Journal Informatic and Information Technology*, vol. 2, pp. 145-156, Agustus 2023.
- [3] B. V. Tarissa and T. D. , “PENERAPAN MACHINE LEARNING DAN DEEP LEARNING PADA PENINGKATAN DETEKSI CREDIT CARD FRAUD-A SYSTEMATIC LITERATURE REVIEW,” *DIPONEGORO JOURNAL OF ACCOUNTING*, vol. 13, pp. 1-15, 2024.
- [4] S. P. and T. D. , “PENERAPAN MACHINE LEARNING, DEEP LEARNING, DAN DATA MINING DALAM DETEKSI KECURANGAN LAPORAN KEUANGAN -A SYSTEMATIC LITERATURE REVIEW,” *DIPONEGORO JOURNAL OF ACCOUNTING*, vol. 13, pp. 1-12, 2024.
- [5] E. P. T. B. A. Hartanto and S. S. , “Penerapan Machine Learning untuk Prediksi Bencana Banjir,” *Jurnal Sistem Informasi Bisnis 01*, vol. 14, pp. 62-76, 2023.
- [6] R. L. and B. W. , “Penerapan Teknik Deep Learning(Long Short Term Memory) dan Pendekatan Klasik (Regresi Linier) dalam Prediksi Pergerakan Saham BRI,” *Jurnal Informatika dan Bisnis*, vol. 12, pp. 42-54., 2023.
- [7] Y. R. A. N. and W. H. , “Analisis Perbandingan Algoritma Optimasi pada Random Forest untuk Klasifikasi Data Bank Marketing,” *JURNAL RESTI*, vol. 5, pp. 187-192, 2021.
- [8] F. I. “Teknik Machine Learning untuk Bank Marketing Dataset,” *Seminar Nasional Informatika (SENATIKA)*, pp. 393-397, 2021.
- [9] W. C. Utomo, “Prediksi Pergerakan Saham BBRI ditengah Issue Ancaman Resesi 2023 dengan Pendekatan Machine Learning,” *Jurnal Teknologi dan Manajemen Informatika*, vol. 9, pp. 20-27, 2023.
- [10] K. H. and A. V. , “Analisis Faktor Tingkat Kebahagiaan Negara Menggunakan Data World Happiness Report dengan Metode Regresi Linier,” *Jurnal Keilmuan dan Aplikasi Teknik Informatika*, vol. 15, pp. 1-7, 2023.
- [11] Y. A. Mustofa, “Analisis Sentimen Terhadap Penggunaan Aplikasi Shopee Menggunakan Algoritma Support Vector Machine (SVM),” *Jambura Journal of Electrical and Electronics Engineering*, vol. 5, pp. 32-35, 2023.
- [12] A. D. D. A. Zayanti, I. R. F. F. Ramadhan and G. , “PERBANDINGAN ALGORITMA SUPPORT VECTOR MACHINE (SVM) DAN,” *Jurnal Kecerdasan Buatan dan Teknologi Informasi*, vol. 4, pp. 33-42, 2025.
- [13] R. E. Ako, F. O. Aghware, M. D. Okpor, M. I. Akazue, R. E. Yoro, A. A. Ojugo, D. R. I. M. Setiadi, C. C. Odiakaose, R. A. Abere, F. . U. Emordi, V. O. Geteloma and P. O. Ejeh, “Effects of Data Resampling on Predicting Customer Churn via a Comparative Tree-based Random Forest and XGBoost,” *Journal of Computing Theories and Applications*, vol. 2, 2024.
- [14] J. and C. F. , “KLASIFIKASI DNA-BINDING PROTEIN MENGGUNAKAN METODE BIDIRECTIONAL GATED RECURRENT UNIT (BIGRU).,” 2023.
- [15] . P. . D. and . R. , “Prediksi Harga Saham Syariah menggunakan Bidirectional Long Short term Memory (BiLSTM) dan algoritma Grid Search,” *JAMBURA JOURNAL OF MATHEMATICS*, vol. 6, pp. 39-45, 2024.
- [16] N. s. and E. N. Putri, “Uji Aktivitas Senyawa 5-hydroxymethylfurfural dari Tanaman Abiu (*Pouteria caimito*) Sebagai Kandidat Obat Kanker Tulang Menggunakan Metode In Silico.,” *Biological Science and Education Journal*, vol. 4, pp. 67-74, 2024.