ANALISIS KINERJA ALGORITMA ENSEMBLE DALAM PREDIKSI PERILAKU PEMBELIAN PELANGGAN

Abdul Latif [1], Siti Khotimatul Wildah [2]
[1] Sistem Informasi Kampus Kota Pontianak, Universitas Bina Sarana Informatika
[2] Teknologi Komputer, Universitas Bina Sarana Informatika
Jalan Kramat Raya No.98, Senen, Jakarta Pusat, Indonesia, 10450, Indonesia
Abdul.bll@bsi.ac.id

ABSTRAK

Prediksi perilaku pembelian pelanggan merupakan salah satu aspek terpenting dalam menentukan efektivitas strategi pemasaran sebuah perusahaan. Dalam dunia bisnis, kemampuan untuk memprediksi pola pembelian pelanggan sangat memengaruhi keputusan strategis, seperti pengelolaan inventaris dan personalisasi penawaran. Selain melakukan prediksi perilaku pembelian, pemilihan model atau algoritma yang digunakan juga menjadi faktor krusial dalam memastikan hasil prediksi yang akurat. Penelitian ini menguji kinerja empat algoritma ensemble, yaitu Gradient Boosting, AdaBoost, XGBoost, dan LightGBM, menggunakan dataset perilaku pembelian pelanggan yang mencakup berbagai atribut relevan. Hasil analisis menunjukkan bahwa algoritma ensemble memberikan performa yang sangat baik dalam memprediksi perilaku pembelian pelanggan. Gradient Boosting memberikan nilai akurasi tertinggi sebesar 92,94% dengan tingkat konsistensi yang baik dengan nilai standar deviasi sebesar 0, 0119, diikuti oleh XGBoost dengan akurasi 92,87% dan standar deviasi sebesar 0,0281, AdaBoost sebesar 92,53% dan standar deviasi sebesar 0, 0137serta LightGBM mencapai akurasi sebesar 92,00% dan standar deviasi sebesar 0, 0296. Gradient Boosting terbukti unggul dalam menghasilkan prediksi yang akurat dan konsisten, sementara XGBoost dan LightGBM menawarkan kecepatan dan efisiensi yang tinggi.

Kata kunci : *Perilaku Pembelian Pelanggan, Ensemble, Gradient Boosting, AdaBoost, XGboost, LightGBM*

1. PENDAHULUAN

Dalam lingkungan yang berorientasi pada data, persaingan pasar semakin ketat. Perusahaan mulai memfokuskan perhatian pada pemasaran yang tepat sasaran untuk menekan biaya, meningkatkan efisiensi pemasaran, dan daya saing [1].

Memahami perilaku pembelian pelanggan merupakan kunci untuk meningkatkan pendapatan dan menjaga keunggulan diantara para pesaing. Berbagai industri bisnis telah mengembangkan kebijakan yang beragam untuk menggali potensi pelanggan dengan memanfaatkan analisis statistik [2].

Untuk mewujudkan hal tersebut, diperlukan analisis mendalam terhadap kebiasaan dan preferensi pelanggan yang dapat membantu perusahaan memahami kebutuhan pasar secara lebih spesifik. Analisis perilaku pelanggan menjadi aspek penting bagi perusahaan dalam mendukung pengambilan keputusan. Dengan menganalisis perilaku pelanggan, pola-pola kebiasaan pelanggan dalam membeli produk atau menggunakan jasa dapat diidentifikasi [3].

Pemahaman tentang perilaku pelanggan di masa depan memiliki peran krusial dalam mendukung pengambilan keputusan strategis, seperti dalam proses manufaktur, perencanaan inventaris di gudang dan titik penjualan, serta pengelolaan sumber daya secara efisien di departemen penjualan dan pemasaran [3].

Selain memberikan wawasan strategis, prediksi perilaku pembelian pelanggan juga membantu meningkatkan pengalaman pelanggan melalui personalisasi yang lebih baik, seperti penawaran produk atau layanan yang sesuai dengan preferensi pelanggan.

Salah satu pendekatan yang sering diterapkan dalam memprediksi perilaku pelanggan adalah pembelajaran mesin. Model pembelajaran mesin merupakan alat yang sangat kuat dan efisien. Pembelajaran mesin merupakan salah satu bidang teknik berbasis data yang tumbuh dengan pesat, menghubungkan ilmu komputer dan statistika, serta menjadi fondasi utama dalam pengembangan kecerdasan buatan dan ilmu data [4].

Teknik pembelajaran mesin yang dapat diterapkan adalah data mining, yaitu sebuah proses sistematis yang digunakan untuk menggali, mengekstraksi pola, atau informasi penting yang tersembunyi dalam kumpulan data yang besar dengan menerapkan berbagai metode dan teknik analisis tertentu [5].

Pemilihan metode atau algoritma yang tepat sangat penting untuk mencapai tujuan yang diinginkan. Oleh karena itu, pemahaman yang jelas mengenai tujuan dan cara memproses data sangat diperlukan agar algoritma yang dipilih dapat memberikan hasil yang optimal. Pendekatan *ensemble learning* dapat dijadikan pilihan utama dalam memecahkan masalah prediksi perilaku pembelian pelanggan. Teknik *ensemble learning* menggabungkan beberapa algoritma *machine learning* untuk menghasilkan prediksi yang lebih akurat dibandingkan dengan menggunakan satu model [6].

Ensemble learning dapat dilakukan secara homogen (algoritma yang sama) atau heterogen (algoritma yang berbeda). Keragaman prediktor dasar meningkatkan akurasi model. Ada empat level dalam metode pembelajaran *ensemble* diantaranya level

pengklasifikasi atau regresi, level data, level fitur, dan level kombinasi [7].

Metode *ensemble learning* telah banyak digunakan seperti pada penelitian mengenai deteksi anomali pada penipuan transaksi keuangan, hasil menunjukan bahwa metode *ensemble learning* secara signifikan meningkatkan performa deteksi penipuan dibandingkan model dasar [8].

Penelitian mengenai klasifikasi keputusan kredit dimana hasil menunjukan algoritma *XGBoost* mampu mencapai 1.0 untuk semua metrik evaluasi [6].

Penelitian lain mengenai prediksi perilaku pelanggan dengan menggunakan algoritma *ensemble learning* yaitu *random forest* dan *XGBoost* telah digunakan dimana hasil penelitian memiliki persentase deviasi rata-rata absolut sebesar 8,27 dan persentase akurasi rata-rata sebesar 91,73 hasil yang sangat baik untuk dataset skala menengah.

Berbagai penelitian sebelumnya telah membuktikan bahwa metode *ensemble* secara signifikan dapat meningkatkan akurasi dan konsistensi dalam berbagai aplikasi, termasuk analisis data keuangan, kredit, dan perilaku konsumen. Dengan demikian penelitian ini akan melakukan implementasi dari metode *ensemble learning*. Penelitian ini termasuk kedalam penelitian klasifikasi dan mengusulkan algoritma seperti *Gradient Boosting, AdaBoost, XGBoost,* dan *LightGBM* untuk memprediksi perilaku pembelian pelanggan. Penelitian ini bertujuan untuk menganalisis kinerja keempat algoritma *ensemble* dalam memprediksi perilaku pembelian pelanggan menggunakan dataset perilaku pembelian pelanggan yang di unduh dari repositoy kaggle. Dataset mencakup berbagai atribut relevan. Evaluasi kinerja dilakukan dengan mengukur metrik akurasi, *precision, recall, f1-score* dan standar deviasi untuk mengidentifikasi algoritma yang paling optimal.

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Penelitian yang dilakukan oleh Shui-xia Chen Xiao-kang Wang, Hong-yu Zhang, Jian-qiang Wang mengenai prediksi pembelian pelanggan dari perspektif data tidak seimbang menggunakan kerangka kerja pembelajaran mesin berbasis algoritma factorization machine. Hasil eksperimen menunjukkan bahwa kerangka kerja yang diusulkan mencapai akurasi tinggi dengan sebagian besar metrik performa hampir atau di atas 90% untuk rata-rata tertimbang dan rata-rata makro pada data uji [9].

Penelitian yang dilakukan oleh Subhatav Dhali, Monalisha Pati, dan Soumi Ghosh membahas mengenai prediksi perilaku pelanggan menggunakan algoritma *ensemble learning*, yaitu *Random Forest* dan *XGBoost*. Hasil penelitian menunjukkan bahwa model tersebut menghasilkan rata-rata persentase deviasi absolut sebesar 8,27 dan rata-rata tingkat akurasi sebesar 91,73%.[10].

Penelitian yang dilakukan oleh Pejman Ebrahimi et al. [11] membahas pengaruh inovasi teknologi startup dan kinerja manajemen hubungan pelanggan terhadap partisipasi pelanggan, penciptaan nilai bersama, dan perilaku pembelian konsumen. Hasil penelitian menunjukkan bahwa inovasi teknologi startup memiliki pengaruh positif dan signifikan terhadap partisipasi pelanggan, sementara kernel polinomial dalam algoritma SVM menunjukkan akurasi tertinggi sebesar 94,6% dalam memverifikasi keakuratan model.

2.2. *Data Mining*

Data mining adalah proses ekstraksi atau penyaringan data dari kumpulan data berukuran besar melalui serangkaian langkah tertentu untuk memperoleh informasi yang bernilai. Proses ini dapat dimanfaatkan oleh perusahaan besar untuk menganalisis data dan memperoleh informasi yang berguna dalam mendukung serta meningkatkan kinerja bisnis perusahaan [12].

Proses ini juga bertujuan untuk menemukan, mengekstraksi pola, atau informasi penting yang tersembunyi dalam kumpulan data yang besar dengan menerapkan berbagai metode dan teknik analisis tertentu [5].

2.3. *Ensemble*

Metode *ensemble* merupakan algoritma dalam pembelajaran mesin yang menggabungkan beberapa model pembelajaran untuk menghasilkan prediksi yang lebih akurat. Metode ini bertujuan untuk meningkatkan kualitas prediksi dibandingkan dengan menggunakan satu model pembelajaran secara terpisah [13].

Salah satu pendekatan dalam metode ini adalah *ensemble classification*, yang menggabungkan hasil prediksi dari berbagai model klasifikasi individu untuk memperoleh hasil yang lebih tepat dan stabil [14].

2.4. *Gradient Boosting*

Gradient boosting adalah algoritma pembelajaran mesin dalam kategori *ensemble learning* yang menggabungkan beberapa model sederhana (*weak learners*) secara bertahap untuk membentuk model yang lebih kuat. Algoritma ini bekerja dengan mengurangi kesalahan prediksi dari model sebelumnya dan fokus pada pola data yang sulit diprediksi. *Gradient boosting* memberi bobot lebih pada data yang sulit dijelaskan oleh model sebelumnya, sehingga meningkatkan akurasi seiring iterasi. Dengan kemampuan menangani data yang kompleks dan hubungan non-linear, *gradient boosting* menjadi pilihan yang efektif untuk analisis klaster dan menghasilkan prediksi yang akurat serta efisien [15].

2.5. *AdaBoost*

Adaptive boosting (*Adaboost*) adalah salah satu jenis algoritma *boosting* yang banyak digunakan dan merupakan metode *ensemble learning* yang bertujuan

meningkatkan kinerja model melalui teknik *boosting* [16].

Adaboost adalah metode dalam *data mining* yang bertujuan untuk meningkatkan akurasi pada teknik klasifikasi. Algoritma ini mudah dikombinasikan dengan berbagai metode klasifikasi lainnya. *Adaboost* bekerja dengan memilih *weak classifiers* melalui proses *feature selection* dan menggabungkannya untuk membentuk sebuah *strong classifier* [17].

2.6. *XGBoost*

Extreme gradient boosting (*XGBoost*) adalah algoritma pembelajaran mesin yang sangat andal dan telah terbukti memberikan akurasi serta performa unggul di berbagai bidang. *XGBoost* beroperasi dengan membangun serangkaian pohon keputusan secara bertahap, di mana setiap pohon berikutnya bertugas memperbaiki kesalahan yang dibuat oleh pohon sebelumnya [18].

2.7. *LightGBM*

LightGBM adalah algoritma pembelajaran mesin yang andal dan efisien untuk tugas klasifikasi dan regresi. Dengan kecepatan tinggi, akurasi yang baik, serta kemudahan dalam penggunaannya, algoritma ini menjadi pilihan yang populer untuk berbagai jenis aplikasi [19].

2.8. Evaluasi Model

Evaluasi dilakukan dengan tujuan untuk mengukur tingkat akurasi hasil klasifikasi, yang mencakup perhitungan nilai-nilai tertentu yang telah ditetapkan sebelumnya. Proses evaluasi ini didasarkan pada analisis matriks klasifikasi (*confusion matrix)* sebagai alat utama untuk menilai kinerja model (Karimah et al., 2024). Selain itu pengukuran dilakukan menggunakan berbagai metrik, termasuk akurasi, presisi, *recall* (sensitivitas), dan *f1-score* [20].

2.8.1. *Confusion Matrix*

Confusion Matrix digunakan untuk mengevaluasi kinerja model klasifikasi. *False Positive* (FP) dan *False Negative* (FN) adalah kesalahan klasifikasi, di mana kelas positif atau negatif terdeteksi salah. Sebaliknya, *True Negative* (TN) menunjukkan kasus di mana kelas positif dan negatif berhasil diklasifikasikan dengan benar [21].

2.8.2. *Precision*

Precision merupakan rasio antara jumlah data positif yang diprediksi dengan benar terhadap total data yang diprediksi sebagai positif, yang mencerminkan perbandingan antara data yang benar dan salah yang diprediksi sebagai positif [22].

$$Precision = \sum \frac{TP}{TP + PF} \quad (1)$$

2.8.3. *Recall*

Recall adalah proporsi data positif yang berhasil diidentifikasi dengan benar dibandingkan dengan total data yang sebenarnya positif, termasuk data yang teridentifikasi dengan benar serta data positif yang tidak terdeteksi (false negative) [22].

$$All\ Recall = \sum \frac{TP}{TP + FN} \quad (2)$$

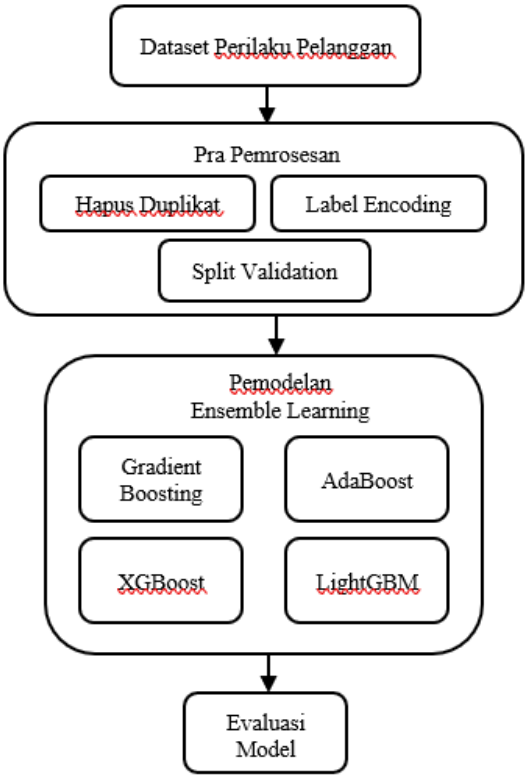
2.8.4. *Accuracy*

Akurasi digunakan untuk menilai kinerja model klasifikasi dengan mengukur tingkat ketepatan metode klasifikasi yang diterapkan [23].

$$Accuracy = \frac{\sum TP}{Total\ Data} x\ 100\% \quad (3)$$

3. METODE PENELITIAN

Penelitian ini bertujuan untuk menganalisis kinerja algoritma *ensemble* dalam memprediksi perilaku pembelian pelanggan, dengan memanfaatkan dataset yang diambil dari *repository Kaggle*. Dalam konteks pengembangan strategi pemasaran yang lebih efektif, prediksi perilaku pelanggan sangat penting untuk meningkatkan efisiensi dan daya saing perusahaan. Tahapan penelitian dapat dilihat pada gambar 1.



Gambar 1. Metode Penelitian

3.1. *Dataset*

Dataset yang digunakan pada penelitian ini berisi informasi terkait perilaku pelanggan dalam melakukan pembelian. *Dataset* ini mencakup 1500 data transaksi

pembelian, yang dapat digunakan untuk menganalisis faktor-faktor yang memengaruhi keputusan pembelian pelanggan. *Dataset* yang digunakan dalam penelitian ini adalah *dataset* asli milik Mr. Rabie El Kharoua yang di publikasikan di *repository Kaggle*. Atribut dapat dilihat pada tabel 1.

Tabel 1. Data Atribut

No	Atribut	Deskripsi
1	*Age*	Usia pelanggan
2	*Gender*	Jenis kelamin pelanggan
3	*Annual Income*	Pendapatan tahunan pelanggan dalam dolar
4	*Number of Purchases*	Total jumlah pembelian yang dilakukan pelanggan
5	*Product Category*	Kategori produk yang dibeli (0: Elektronik, 1: Pakaian, 2: Barang Rumah Tangga, 3: Kecantikan, 4: Olahraga)
6	*Time Spent on Website*	Waktu yang dihabiskan pelanggan di situs web dalam menit
7	*Loyalty Program*	Keanggotaan program loyalitas
8	*Discounts Availed*	Jumlah diskon yang dimanfaatkan oleh pelanggan (rentang: 0-5)
9	*PurchaseStatus* (Target Variable)	Kemungkinan pelanggan melakukan pembelian

Target variabel yaitu atribut *PurchaseStatus* yang menunjukkan kemungkinan pelanggan untuk melakukan pembelian, dengan dua kategori utama:

a. 0 (*No Purchase*): Pelanggan tidak melakukan pembelian, yang mencakup 48% dari total data.
b. 1 (*Purchase*): Pelanggan melakukan pembelian, yang mencakup 52% dari total data.

Distribusi ini menunjukkan bahwa data relatif seimbang antara pelanggan yang melakukan pembelian dan yang tidak.

3.2. Pra Proses

Tahapan praproses data penting dilakukan dalam analisis data dan pembelajaran mesin untuk mempersiapkan data mentah menjadi bentuk yang siap digunakan untuk pemodelan. Adapun langkah-langkah yang dilakukan adalah sebagai berikut:

3.2.1. Hapus Duplikat

Pada tahap ini, dilakukan penghapusan data duplikat yang teridentifikasi dalam dataset. Dari total 1500 data, ditemukan sebanyak 112 data duplikat. Oleh karena itu, data duplikat tersebut dihapus sehingga jumlah data yang tersisa menjadi 1388 data unik. Proses ini bertujuan untuk memastikan integritas data dan mencegah pengaruh negatif pada hasil analisis akibat redundansi informasi.

3.2.2. *Label Encoding*

Label encoding dilakukan untuk mengubah variabel kategorikal menjadi bentuk numerik. Transformasi ini bertujuan agar variabel kategorikal dapat digunakan dalam model pembelajaran mesin, di mana setiap kategori dalam variabel tersebut diberi label numerik yang unik.

3.2.3. *Split Validation*

Pembagian data dilakukan dengan memisahkan *dataset* menjadi dua bagian, yaitu 80% untuk pelatihan dan 20% untuk pengujian. Langkah ini bertujuan untuk memberikan sebagian besar data kepada model untuk proses pelatihan, sementara sisanya digunakan untuk menilai kinerja model pada data baru yang belum pernah diakses sebelumnya.

3.3. Pemodelan *Ensemble*

Pada penelitian ini, dilakukan tahap pemodelan dengan memanfaatkan metode *ensemble* untuk meningkatkan performa prediksi. Metode *ensemble* yang digunakan meliputi *Gradient Boosting, AdaBoost, XGBoost,* dan *LightGBM*. Pemilihan model-model ini didasarkan pada kemampuannya dalam menghasilkan prediksi yang akurat dengan memadukan kontribusi dari sejumlah model sederhana (*weak learners*).

3.4. Evaluasi Model

Evaluasi model dilakukan untuk menilai performa hasil pengujian algoritma *Ensemble* yang digunakan meliputi *Gradient Boosting, AdaBoost, XGBoost,* dan *LightGBM*. Metrik yang digunakan untuk mengukur kinerja model meliputi akurasi, presisi, *recall*, *AUC*, *f1-score* dan *confusion matrix*.

4. HASIL DAN PEMBAHASAN

Dataset dievaluasi dengan membandingkan kinerja beberapa algoritma pembelajaran mesin, yaitu *Gradient Boosting*, *AdaBoost*, *XGBoost*, dan *LightGBM*. Untuk mengevaluasi performa model, data dipisahkan menjadi dua subset, yaitu data pelatihan dan data pengujian. Model dilatih menggunakan subset data pelatihan dan divalidasi menggunakan subset data pengujian.

Penelitian ini menggunakan *cross-validation* untuk mengevaluasi performa model secara objektif dan mencegah *overfitting* maupun *underfitting*. Data dibagi menjadi beberapa subset (*fold*), di mana model dilatih pada k-1 subset dan diuji pada subset sisanya. Proses ini diulang beberapa kali, dan rata-rata akurasi dari seluruh iterasi dihitung untuk mendapatkan hasil akhir. *Cross-validation* memastikan evaluasi model tidak terpengaruh oleh pembagian data tertentu dan memberikan gambaran yang konsisten tentang kemampuan model memprediksi data baru [24].

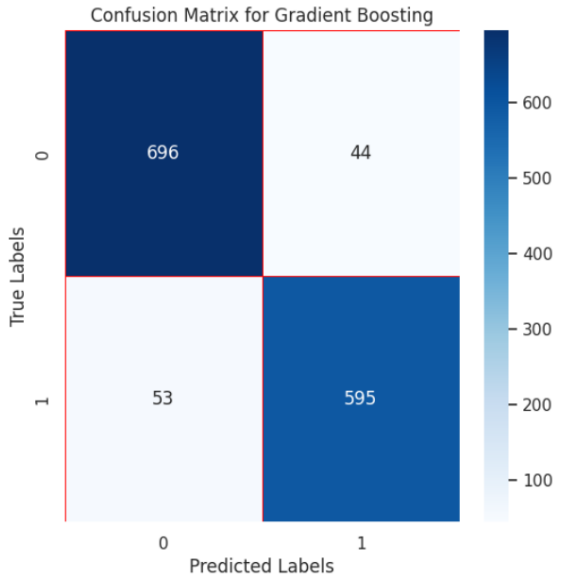
Dalam penelitian ini, digunakan *10-fold cross-validation* untuk membandingkan performa beberapa algoritma *ensemble*, berdasarkan rata-rata akurasi dan standar deviasi. Hasil evaluasi dari metode *ensemble* yang digunakan ditampilkan pada Tabel 2 berikut.

Tabel 2. Hasil Evaluasi Model *Ensemble*

No	Metode	Akurasi	Standar Deviasi
1	*Gradient Boosting*	92,94%	0.0119
2	*XGBoost*	92,87%	0.0281
3	*AdaBoost*	92,53%	0.0137
4	*LightGBM*	92,00%	0.0296

Tabel di atas menunjukkan bahwa *gradient boosting* memiliki rata-rata akurasi tertinggi dibandingkan algoritma lainnya, dengan performa yang stabil sebagaimana ditunjukkan oleh nilai standar deviasinya yang rendah. Hal ini menjadikan *gradient boosting* sebagai algoritma dengan kinerja terbaik dalam penelitian ini.

Setelah *gradient boosting* diidentifikasi sebagai algoritma dengan performa terbaik berdasarkan hasil *cross-validation*, model ini dievaluasi lebih lanjut menggunakan *confusion matrix*. Hasil pengujian *confusion matrix* dapat dilihat pada Gambar 2.



Gambar 2. *Confusion Matrix*

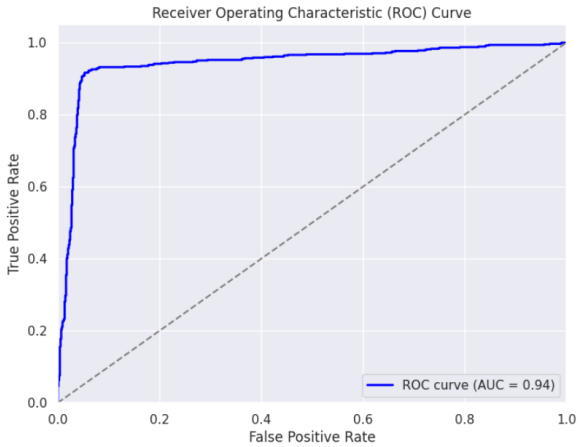
Berdasarkan evaluasi menggunakan *confusion matrix*, model *gradient boosting* menunjukkan performa yang baik. Terdapat 696 data berhasil diprediksi dengan benar sebagai tidak melakukan pembelian (*True Negatives*), dan 595 data diprediksi dengan benar sebagai melakukan pembelian (*True Positives*). Kesalahan pelabelan relatif kecil, yaitu 44 data tidak melakukan pembelian yang salah diprediksi sebagai melakukan pembelian (*False Positives*), dan 53 data melakukan pembelian yang salah diprediksi sebagai tidak melakukan pembelian (*False Negatives*).

Hasil ini menunjukkan bahwa model mampu memprediksi data dengan akurat untuk kedua kelas. Evaluasi ini juga menjadi dasar untuk menghitung metrik performa lainnya. Berikut laporan klasifikasi yang dapat dilihat pada Tabel 3.

Tabel 3. Laporan Klasifikasi

Label	Precision	Recall	F1-Score	Support (Data)
0: Tidak	93%	94%	93%	740
1: Ya	93%	92%	92%	648
Accuracy	-	-	**93%**	**1388**
Macro Avg	93%	93%	93%	1388
Weighted Avg	93%	93%	93%	1388

Model *gradient boosting* menunjukkan kinerja yang sangat baik dalam memprediksi kemungkinan pelanggan melakukan pembelian dengan akurasi keseluruhan sebesar 93%. *Precision*, *recall*, dan *f1-score* pada kedua label juga tinggi dan konsisten, yaitu masing-masing sebesar 93% untuk pelanggan yang tidak melakukan pembelian dan 92-93% untuk pelanggan yang melakukan pembelian. Hal ini menunjukkan bahwa model mampu mengklasifikasikan kedua kategori secara seimbang dan akurat, menjadikannya andal dalam mendukung analisis perilaku pelanggan. Evaluasi lain ditunjukan dengan analisi kurva ROC yang dapat dilihat pada Gambar 3.



Gambar 3. Kurva ROC

Kurva ROC (*Receiver Operating Characteristic*) menunjukkan bahwa model *gradient boosting* memiliki kinerja yang sangat baik dalam memprediksi perilaku pembelian pelanggan dengan nilai AUC sebesar 0,94, model ini mampu membuat prediksi yang akurat.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian, dapat disimpulkan bahwa dalam analisis perilaku pembelian pelanggan menggunakan algoritma *gradient boosting*, diperoleh hasil yang sangat baik dengan akurasi sebesar 92,94%, *recall* 94%, *precision* 93%, *f1-score* 93%, dan AUC 0,94. Hasil ini menunjukkan bahwa model *gradient boosting* mampu mengklasifikasikan data dengan akuray, baik untuk pelanggan yang tidak melakukan pembelian maupun pelanggan yang melakukan pembelian. Dibandingkan dengan algoritma *ensemble*

lainnya, seperti *XGBoost*, *AdaBoost*, dan *LightGBM*, *gradient boosting* menunjukkan kinerja yang lebih stabil dan akurat, dengan standar deviasi yang rendah. Evaluasi menggunakan *confusion matrix* juga mendukung hasil ini, dengan kesalahan klasifikasi yang relatif kecil. Penelitian selanjutnya dapat dilakukan dengan membandingkan *gradient boosting* dengan algoritma lain yang lebih beragam, serta menerapkan praproses data yang lebih mendalam.

DAFTAR PUSTAKA

[1] J. Li, S. Pan, L. Huang, and X. Zhu, "A machine learning based method for customer behavior prediction," *Teh. Vjesn.*, vol. 26, no. 6, pp. 1670–1676, 2019, doi: 10.17559/TV-20190603165825.

[2] A. M. Choudhury and K. Nur, "A machine learning approach to identify potential customer based on purchase behavior," *1st Int. Conf. Robot. Electr. Signal Process. Tech. ICREST 2019*, no. 1, pp. 242–247, 2019, doi: 10.1109/ICREST.2019.8644458.

[3] Jeffry, S. Usman, and F. Aziz, "Analisis Perilaku Pelanggan menggunakan Metode Ensemble Logistic Regression," *J. Teknol. Dan Ilmu Komput. Prima*, vol. 6, no. 2, pp. 90–97, 2023, doi: 10.34012/jutikomp.v6i2.4258.

[4] L. Zhang *et al.*, "A review of machine learning in building load prediction," *Appl. Energy*, vol. 285, no. July 2020, p. 116452, 2021, doi: 10.1016/j.apenergy.2021.116452.

[5] S. K. Wildah, S. Agustiani, M. R. R. S, W. Gata, and H. M. Nawawi, "Deteksi Penyakit Alzheimer Menggunakan Algoritma Naïve Bayes Dan Correlation Based Feature Selection," *J. Inform.*, vol. 7, no. 2, pp. 166–173, 2020, doi: 10.31294/ji.v7i2.8226.

[6] Jan Melvin Ayu Soraya Dachi and Pardomuan Sitompul, "Analisis Perbandingan Algoritma XGBoost dan Algoritma Random Forest Ensemble Learning pada Klasifikasi Keputusan Kredit," *J. Ris. Rumpun Mat. Dan Ilmu Pengetah. Alam*, vol. 2, no. 2, pp. 87–103, 2023, doi: 10.55606/jurrimipa.v2i2.1470.

[7] S. S. Nusrhendratno, "Sintesis Fitur Density Based Feature Selection (DBFS) dan AdaBoots dengan XGBoost Untuk Meningkatkan Performa Model Prediksi," *Pros. Sains Nas. dan Teknol.*, vol. 12, no. 1, p. 305, 2022, doi: 10.36499/psnst.v12i1.6997.

[8] D. R. K. Saputra, Y. V. Via, and A. N. Sihananto, "Deteksi Anomali Menggunakan Ensemble Learning Dan Random Oversampling Pada Penipuan Transaksi Keuangan," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 3, pp. 2779–2788, 2024, doi: 10.23960/jitet.v12i3.4910.

[9] S. xia Chen, X. kang Wang, H. yu Zhang, and J. qiang Wang, "Customer purchase prediction from the perspective of imbalanced data: A machine learning framework based on factorization machine," *Expert Syst. Appl.*, vol. 173, no. February, p. 114756, 2021, doi: 10.1016/j.eswa.2021.114756.

[10] S. Dhali, M. Pati, S. Ghosh, and C. Banerjee, "An Efficient Predictive Analysis Model of Customer Purchase Behavior using Random Forest and XGBoost Algorithm," *2020 IEEE Int. Conf. Converg. Eng. ICCE 2020 - Proc.*, pp. 416–421, 2020, doi: 10.1109/ICCE50343.2020.9290576.

[11] P. Ebrahimi, A. Salamzadeh, M. Soleimani, S. M. Khansari, H. Zarea, and M. Fekete-Farkas, "Startups and Consumer Purchase Behavior: Application of Support Vector Machine Algorithm," *Big Data Cogn. Comput.*, vol. 6, no. 2, 2022, doi: 10.3390/bdcc6020034.

[12] M. S. Azad, S. S. Khan, R. Hossain, R. Rahman, and S. Momen, "Predictive modeling of consumer purchase behavior on social media: Integrating theory of planned behavior and machine learning for actionable insights," *PLoS One*, vol. 18, no. 12 December, pp. 1–26, 2023, doi: 10.1371/journal.pone.0296336.

[13] E. Mardiani *et al.*, "Analisis Prediksi Pendapatan Penduduk dengan Metode K-Nearest Neighbor, Decision Tree, Naive Bayes, Ensemble Methods, dan Linear Regression," *J. Soc. Sci. Res.*, vol. 3, no. 4, pp. 8667–8679, 2023.

[14] Steven Joses, D. Yulvida, and S. Rochimah, "Pendekatan Metode Ensemble Learning untuk Prakiraan Cuaca menggunakan Soft Voting Classifier," *J. Appl. Comput. Sci. Technol.*, vol. 5, no. 1, pp. 72–80, 2024, doi: 10.52158/jacost.v5i1.741.

[15] F. Septian, "Optimasi Klusterisasi pada Lama Tempo Pekerjaan Berbasis Gradient Boost Algorithm," 2023.

[16] M. Siddik Hasibuan and H. Harahap, "Penerapan Metode Haar-Like Feature Dan Algoritma Adaboost Dalam Penentuan Klasifikasi Hama Tanaman Kopi," *J. Sci. Soc. Res.*, vol. 4307, no. 1, pp. 87–93, 2024, [Online]. Available: http://jurnal.goretanpena.com/index.php/JSSR

[17] R. T. Febianto, D. Suranti, and R. T. Alinse, "Penerapan Algoritma Adaboost Dalam Mengetahui Pola Pengguna Kb Di Puskesmas Tanjung Harapan," *J. Sci. Soc. Res.*, vol. 4307, no. 1, pp. 145–155, 2024, [Online]. Available: http://jurnal.goretanpena.com/index.php/JSSR

[18] K. Aditya, A. Wisnu, and A. M. A. Rahim, "Analisis Perbandingan Algoritma XGBoost Dan Algoritma Random Forest Untuk Klasifikasi Data Kesehatan Mental," vol. 2, no. 5, pp. 808–818, 2024.

[19] F. Duran, F. Wijaya, Y. R. Hulu, M. Harahap, and A. Prabowo, "Perbandingan Kinerja Algoritma Random Forest Classifier Dan Lightgbm Classifier Untuk Prediksi Penyakit Jantung," *Data Sci. Indones.*, vol. 3, no. 2, pp. 98–103, 2024, doi: 10.47709/dsi.v3i2.3831.

[20] F. Aziz, P. Ishak, and S. Abasa, "Klasifikasi

Depresi Menggunakan Support Vector Machine: Pendekatan Berbasis Data Text Mining," *J. Pharm. Appl. Comput. Sci.*, vol. 2, no. 2, pp. 33–38, 2024, doi: 10.59823/jopacs.v2i2.53.

[21] W. Priatna, "Dampak Pengambilan Sampel Data untuk Optimalisasi Data tidak seimbang pada Klasifikasi Penipuan Transaksi E-Commerce," *Indones. J. Comput. Sci.*, vol. 13, no. 2, pp. 3070–3079, 2024, doi: 10.33022/ijcs.v13i2.3698.

[22] N. Arifin, U. Enri, and N. Sulistiyowati, "Penerapan Algoritma Support Vector Machine (SVM) dengan TF-IDF N-Gram untuk Text Classification," *STRING (Satuan Tulisan Ris. dan Inov. Teknol.*, vol. 6, no. 2, p. 129, 2021, doi: 10.30998/string.v6i2.10133.

[23] D. Alita, Y. Fernando, and H. Sulistiani, "Implementasi Algoritma Multiclass Svm Pada Opini Publik Berbahasa Indonesia Di Twitter," *J. Tekno Kompak*, vol. 14, no. 2, p. 86, 2020, doi: 10.33365/jtk.v14i2.792.

[24] T. M. Innayah and I. K. D. Nuryana, "Model Klasifikasi Produk Paling Diminati Pada Penjualan Sandal Wanita Menggunakan Metode Naïve Bayes (Studi Kasus pada UMKM Ann- D ' Mello Sandals Krian Sidoarjo)," vol. 05, no. 03, pp. 129–138, 2024.