JOISIE (Journal Of Information Systems And Informatics Engineering)
Vol. 9, No.1, Juni 2025, Hlm 100-108
Received: 14 April 2025 Revised: 5 Mei 2025, Accepted: 13 Mei 2025

p- ISSN: 2503-5304
e- ISSN: 2527-3116

ANALISIS KOMPARATIF MODEL *DATA MINING* DALAM PREDIKSI KETEPATAN PENYELESAIAN *SERVICE LEVEL AGREEMENT*

Aisyah Fatimah[1)], Ken Ditha Tania[2)], Allsela Meiriza[3)]
[1-3] Program Studi Sistem Informasi, Universitas Sriwijaya, Jalan Palembang-Prabumulih, KM 32 Inderalaya,Kabupaten Ogan Ilir, Sumatera Selatan (30662)
email : 2aisyahfatimah@gmail.com, kenya.tania@gmail.com, allsela@unsri.ac.id

Abstract

Compliance with the Service Level Agreement (SLA) is crucial in IT service management to ensure operational continuity and service quality by setting clear expectations for incident resolution time. However, many organizations struggle to predict whether incident tickets will meet SLA targets, leading to user dissatisfaction, issue escalation, and increased workload for IT teams. The large volume of daily tickets makes it difficult to manually identify and prioritize high-risk cases. This study aims to develop a machine learning classification model to predict whether a ticket will meet its SLA. Four algorithms were evaluated: XGBoost, Random Forest, Decision Tree, and Logistic Regression. The modeling process involved data preprocessing, categorical feature encoding, feature selection using Random Forest, class balancing with SMOTE, and hyperparameter tuning using Optuna. Results show that XGBoost outperforms other models with an accuracy of 99.98%, precision of 0.9437, recall of 0.9710, and F1-score of 0.9571. Beyond high accuracy and efficiency, the model excels in interpretability using SHAP (SHapley Additive exPlanations), which reveals feature contributions to predictions. In conclusion, XGBoost is recommended as the most reliable model to serve as a strategic tool for IT service managers in identifying incidents at high risk of failing to meet the SLA.
Keywords: *Machine Learning, Imbalanced Data, SMOTE, SHAP, OPTUNA*

Abstrak

Kepatuhan terhadap *Service Level Agreement* (SLA) sangat penting dalam manajemen layanan teknologi informasi untuk menjamin kualitas layanan dan mengatur ekspektasi penyelesaian insiden. Namun, banyak organisasi kesulitan memprediksi apakah tiket insiden akan memenuhi SLA, yang dapat menyebabkan ketidakpuasan pengguna, eskalasi masalah, dan beban kerja tinggi bagi tim IT. Tingginya volume tiket harian membuat identifikasi manual terhadap tiket berisiko tinggi menjadi tidak efektif. Penelitian ini bertujuan mengembangkan model klasifikasi berbasis *machine learning* untuk memprediksi kepatuhan tiket terhadap SLA. Empat algoritma dievaluasi: XGBoost, Random Forest, Decision Tree, dan Logistic Regression. Tahapan mencakup *preprocessing*, *encoding* fitur kategorikal, seleksi fitur berbasis Random Forest, penyeimbangan data menggunakan SMOTE, dan *hyperparameter tuning* dengan Optuna. Hasil menunjukkan XGBoost memiliki performa terbaik dengan akurasi 99,98%, *precision* 0,9437, *recall* 0,9710, dan F1-*score* 0,9571. Selain akurat dan efisien, model ini unggul secara interpretatif melalui SHAP, yang menjelaskan kontribusi tiap fitur. Kesimpulannya, XGBoost direkomendasikan sebagai model paling andal untuk menjadi alat bantu strategis bagi manajer layanan TI dalam mengidentifikasi insiden yang berisiko tinggi gagal memenuhi SLA.
Kata Kunci: *Machine Learning*, *Imbalanced Data, SMOTE, SHAP, OPTUNA*

1. PENDAHULUAN

Service Level Agreement atau Perjanjian Tingkat Layanan adalah dokumen kontraktual yang memuat komitmen penyedia layanan kepada pengguna terkait kualitas layanan, mencakup aspek seperti waktu tanggap, stabilitas sistem, dan ketersediaan (Suja & Booba, 2020). Pada umumnya, SLA tradisional mencakup beberapa indikator kinerja seperti kapasitas bandwidth, tingkat ketersediaan layanan, waktu tunda (*latency*), tingkat

https://doi.org/10.35145/joisie.v9i1.4943

kehilangan paket data, serta frekuensi kesalahan. Selain itu, SLA juga mengukur kecepatan respons terhadap tiket layanan yang masuk melalui aplikasi help desk (seperti *call center*), serta menetapkan waktu minimal sebelum permintaan jadwal pemeliharaan dapat diajukan (Gonçalves et al., 2020; Nugraha et al., 2018; Amazonas et al., 2019). SLA juga mencantumkan ketentuan sanksi atau kompensasi apabila penyedia layanan gagal memenuhi standar yang telah disepakati (Supono, 2020). Lebih dari sekadar alat pengatur hubungan antara penyedia dan pengguna layanan, penerapan SLA juga dapat berfungsi sebagai indikator kinerja utama *(Key Performance Indicator*/KPI) bagi layanan internal. Hal ini mencerminkan kinerja perusahaan secara keseluruhan serta memberikan gambaran mengenai pengalaman pengguna terhadap layanan yang diberikan (Suja & Booba, 2020).

Dalam berbagai studi, model berbasis pohon keputusan, khususnya *Extreme Gradient Boosting* (XGBoost), menunjukkan kinerja prediktif yang unggul dibandingkan metode lainnya. Misalnya, dalam prediksi output daya fotovoltaik, XGBoost menghasilkan nilai *Mean Square Error* (MSE) terendah dan *regression value* tertinggi, melampaui model-model lainnya (Didavi et al., 2021). Dalam konteks prediksi risiko asuransi, algoritma ini kembali menunjukkan performa terbaik dengan nilai *Area Under Curve* (AUC) dan F1-*score* tertinggi (Sahai et al., 2023). Dalam prediksi waktu penyelesaian layanan pelanggan, XGBoost juga terbukti lebih unggul dibandingkan teknik lainnya, dengan skor galat yang rendah dan tingkat akurasi yang tinggi (Tai et al., 2023). Sementara itu, studi yang dilakukan oleh Ansori et al., (2023) menunjukkan bahwa algoritma Random Forest mampu memprediksi waktu tempuh *Service Level Agreement* (SLA) pada pengiriman PT. Pos Indonesia dengan akurasi sebesar 83,86%. Teknik seperti SHAP (*SHapley Additive exPlanations*) dan *Feature Importance* digunakan untuk mengidentifikasi kontribusi masing-masing variabel terhadap hasil prediksi, sehingga hasil analisis menjadi lebih transparan dan dapat dipertanggungjawabkan (Ergün, 2023). Tidak hanya dari sisi akurasi, studi-studi tersebut juga menyoroti pentingnya interpretabilitas model, yang kini menjadi perhatian utama dalam implementasi *machine learning*, khususnya di lingkungan bisnis yang membutuhkan dasar pengambilan keputusan yang jelas.

Meskipun telah banyak studi yang membandingkan performa berbagai algoritma *machine learning* dalam konteks prediksi, termasuk dalam ranah *Service Level Agreement* (SLA), sebagian besar masih berfokus pada aspek akurasi semata tanpa memberikan perhatian yang khusus terhadap interpretabilitas model dan relevansi variabel input dalam konteks bisnis (Tan et al., 2022). Banyak penelitian yang menilai kinerja model secara kuantitatif menggunakan metrik seperti akurasi, F1-*score*, atau AUC, namun belum secara mendalam mengevaluasi kontribusi masing-masing fitur terhadap hasil prediksi. Padahal, dalam praktik bisnis, pemahaman terhadap faktor-faktor penyebab keberhasilan atau kegagalan SLA menjadi elemen kunci dalam pengambilan keputusan manajerial dan perbaikan kualitas layanan (Bhargavi & Sekhar, 2021).

Studi ini bertujuan membangun model klasifikasi untuk memprediksi apakah insiden akan diselesaikan sebelum atau setelah tenggat waktu SLA, menggunakan pendekatan *machine learning* dengan *pipeline* yang meliputi *preprocessing, feature selection*, penanganan ketidakseimbangan kelas (SMOTE), *scaling*, *tuning model* dengan Optuna, serta evaluasi performa model secara komprehensif. Empat algoritma digunakan dalam penelitian ini, yakni Decision Tree, Logistic Regression, Random Forest, dan XGBoost. Decision Tree memudahkan visualisasi struktur keputusan dan identifikasi variabel yang paling berpengaruh, sementara Logistic Regression efektif dalam menangani klasifikasi biner serta menyediakan estimasi probabilistik terhadap kemungkinan terjadinya kegagalan SLA (Nur et al., 2023). Di sisi lain, Random Forest dan XGBoost sebagai algoritma *ensemble learning* dinilai unggul dalam meningkatkan akurasi prediksi serta mengurangi *risiko overfitting* (Dachi & Sitompul, 2023). Studi oleh Hou et al., (2022) mendukung keunggulan kedua model ini, di mana Random Forest dan *Gradient Boosting Decision Trees* (GBDT) terbukti melampaui metode tradisional berbasis teori antrian dalam memprediksi performa layanan *call center*, dengan peningkatan performa sebesar 6–9% berdasarkan metrik MAE dan MAPE. Selain itu, XGBoost dikenal luas karena efisiensinya dalam komputasi, skalabilitasnya dalam menangani data berukuran besar, serta kemampuannya dalam mengelola variabel input yang kompleks dan multivariat.

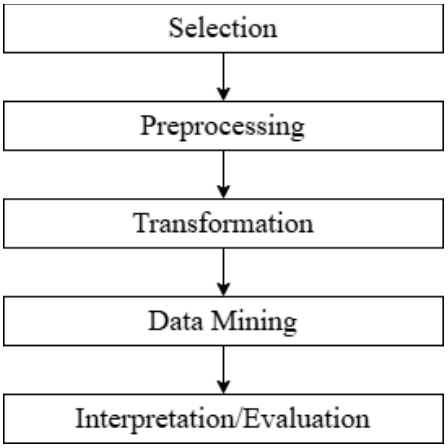
Dataset yang digunakan dalam penelitian ini berasal dari situs sumber terbuka Kaggle dengan judul *"incident_event_log"* (Amaral et al., 2018). Dataset tersebut terdiri atas 24.918 baris dan 34 atribut yang merepresentasikan rangkaian aktivitas dalam proses penanganan insiden pada platform ServiceNow™ yang diimplementasikan oleh sebuah perusahaan di bidang teknologi informasi. Informasi pada log ini telah dilengkapi dengan data tambahan yang diambil dari basis data relasional terkait, serta telah melalui proses

https://doi.org/10.35145/joisie.v9i1.4943

anonimisasi untuk menjaga kerahasiaan dan privasi. Kompleksitas dataset yang mencakup beragam atribut serta adanya ketidakseimbangan kelas menjadikannya sebagai ssu riset yang membutuhkan pendekatan mendalam dan memiliki nilai kebaruan, mengingat aspek tersebut belum banyak dikaji secara mendalam dalam studi sebelumnya.

Berbeda dari penelitian sebelumnya, studi ini menekankan pada aspek interpretabilitas model melalui teknik SHAP untuk menjembatani kebutuhan prediksi yang akurat dan pengambilan keputusan bisnis yang berbasis data. Keterbaharuan lain dari penelitian ini terletak pada penggunaan data nyata dari industri dengan tingkat ketidakseimbangan kelas yang tinggi dan kompleksitas fitur yang mencerminkan kondisi operasional layanan secara aktual. Selain fokus pada akurasi, penelitian ini juga mengevaluasi relevansi variabel input terhadap hasil prediksi untuk menghasilkan insight yang bermakna dalam konteks bisnis, serta mendukung strategi peningkatan kualitas layanan dan kepuasan pelanggan secara berkelanjutan di berbagai sektor industri.

2. METODE PENELITIAN



Gambar 1. Tahapan Penelitian

Penelitian ini menggunakan pendekatan *Knowledge Discovery in Databases* (KDD) yang terdiri dari lima tahapan utama seperti pada Gambar 1. yakni seleksi data, pra proses data, transformasi, *data mining*, dan interpretasi/evaluasi. Tahapan seleksi data bertujuan untuk memilih data yang relevan dari kumpulan data yang tersedia sesuai dengan tujuan analisis (Dewi et al., 2022). Selanjutnya, tahap pra-proses data dilakukan untuk membersihkan data dari nilai yang hilang, data duplikat, maupun inkonsistensi agar data siap digunakan. Tahap transformasi berfokus pada pengubahan data ke dalam format yang sesuai untuk proses *data mining*, seperti normalisasi atau pengkodean ulang variabel. Pada tahap *data mining*, dilakukan proses penggalian pola atau pengetahuan dari data menggunakan algoritma tertentu sesuai dengan kebutuhan penelitian. Terakhir, tahap interpretasi dan evaluasi digunakan untuk menafsirkan hasil *data mining* dan menilai sejauh mana hasil tersebut bermanfaat serta valid terhadap tujuan yang ingin dicapai dalam penelitian (Meneses & Grinstein, 1998).

2.1 Seleksi Data

Dataset yang digunakan dalam penelitian ini diambil dari situs open source *Kaggle*, dengan judul "incident_event_log" (Amaral et al., 2018). Dataset memuat 24.918 baris dan 34 kolom. Dataset ini memuat catatan peristiwa yang menggambarkan jalannya proses manajemen insiden, berdasarkan data hasil pencatatan aktivitas sistem audit pada platform ServiceNow™ yang diterapkan oleh sebuah perusahaan di bidang teknologi informasi. Data dalam log ini telah diperluas dengan informasi tambahan yang diambil dari basis data relasional yang terhubung dengan sistem informasi berbasis proses. Seluruh informasi yang bersifat sensitif telah disamarkan demi menjaga kerahasiaan dan privasi data.

https://doi.org/10.35145/joisie.v9i1.4943

Atribut ‘number’ merupakan pengidentifikasi unik dari setiap insiden yang ada. Karena sifatnya hanya sebagai ID dan tidak memberikan informasi prediktif, atribut ini dibuang dari proses analisis. ‘incident_state’ menggambarkan status tahapan insiden dalam proses manajemen dari awal hingga selesai, sehingga atribut ini dipilih karena penting untuk memahami progres penanganan insiden. ‘active’ adalah atribut boolean yang menunjukkan apakah insiden masih terbuka atau sudah ditutup, dan atribut ini dipertahankan karena berhubungan dengan status penyelesaian.

Selanjutnya, ‘reassignment_count’ menunjukkan seberapa sering insiden berpindah tangan antar tim atau individu, dan dipilih karena berpotensi menunjukkan kompleksitas penyelesaian. Atribut ‘reopen_count’ yang menunjukkan berapa kali insiden dibuka kembali karena solusi tidak memuaskan juga dipertahankan karena berkaitan dengan kualitas penyelesaian. ‘sys_mod_count’ yang menunjukkan jumlah pembaruan insiden juga dipilih karena merepresentasikan dinamika penanganan kasus. Atribut ‘caller_id’, ‘opened_by’, dan ‘sys_created_by’ semuanya adalah identitas pengguna dan dianggap terlalu spesifik serta tidak dapat digeneralisasi, sehingga dibuang. Demikian pula dengan atribut waktu seperti ‘opened_at’, ‘sys_created_at’, dan ‘sys_updated_at’ yang tidak diproses lebih lanjut menjadi fitur waktu yang bermakna, sehingga dibuang.

Atribut ‘contact_type’, yang menunjukkan media pelaporan insiden seperti telepon atau web, dipertahankan karena berpengaruh terhadap waktu tanggap. ‘Location’ juga dipilih karena bisa diolah menjadi fitur kontekstual yang berguna. ‘category’ dan ‘subcategory’, yang menggambarkan tingkat pertama dan kedua layanan terdampak, dipilih karena membantu dalam mengelompokkan jenis insiden. Sementara itu, ‘u_symptom’ berisi deskripsi teks bebas dari pengguna dan dibuang karena membutuhkan pemrosesan teks yang lebih kompleks.

Atribut ‘cmdb_ci’, meskipun mengacu pada item yang terdampak, dibuang karena terlalu spesifik dan variatif. Sebaliknya, atribut ‘impact’ dan ‘urgency’ dipilih karena merupakan input utama dalam menentukan prioritas insiden, yang juga ikut dipertahankan karena penting dalam pengambilan keputusan SLA. Atribut ‘assignment_group’, meskipun berupa ID, tetap dipilih karena umumnya mencerminkan divisi yang cukup stabil dan bisa menjadi indikator penyelesaian. Namun, ‘assigned_to’ dibuang karena terlalu personal.

Atribut ‘knowledge’ menunjukkan apakah basis pengetahuan digunakan dalam penyelesaian dan dipertahankan karena menunjukkan efektivitas resolusi. ‘u_priority_confirmation’ yang menunjukkan validasi prioritas juga dipilih karena meningkatkan keandalan data. Atribut ‘notify’, yang menunjukkan apakah notifikasi dikirim, juga dipertahankan sebagai indikator komunikasi otomatis dalam sistem.

Selanjutnya, atribut seperti ‘problem_id’, ‘rfc’, ‘vendor’, dan ‘caused_by’ dibuang karena semuanya adalah ID yang spesifik dan tidak memberikan informasi umum yang berguna. Atribut ‘Close_code’ yang menjelaskan jenis penyelesaian, tetap dipertahankan karena dapat memberikan konteks tambahan pada hasil insiden. ‘resolved_by’ dibuang karena mengandung informasi individual. Terakhir, ‘resolved_at’ dan ‘closed_at’, meskipun penting, hanya digunakan untuk membentuk label target prediksi (‘sla_met’), dan setelah itu dibuang dari fitur input. Dengan demikian, hanya atribut yang bersifat umum, prediktif, dan tidak menyebabkan kebocoran data yang dipertahankan untuk proses analisis dan prediksi SLA.

2.2 Pra Proses Data

Kolom waktu seperti ‘resolved_at’ dan ‘closed_at’ dikonversi menjadi format *datetime* agar dapat dihitung durasinya. Data dengan nilai kosong pada kedua kolom tersebut dibuang karena tidak dapat digunakan untuk menentukan SLA. Label target baru bernama ‘sla_met’ dibuat, dengan nilai 1 jika selisih antara ‘closed_at’ dan ‘resolved_at’ kurang dari atau sama dengan nol (berarti SLA terpenuhi), dan 0 jika tidak. Atribut target ini menjadi acuan dalam klasifikasi. Kemudian, seluruh fitur kategorikal dikonversi menjadi numerik menggunakan teknik *Label Encoding*. Untuk memastikan tidak ada data hilang yang dapat mengganggu proses *training*, semua nilai kosong pada fitur numerik diisi dengan nilai median dari kolom masing-masing (Yudistira & Putra, 2021).

2.3 Transformasi Data

Pada tahap Transformasi data, data yang telah dibersihkan selanjutnya diubah ke dalam bentuk yang sesuai untuk proses pelatihan model. Proses ini diawali dengan pemilihan fitur penting menggunakan model Random Forest untuk mengidentifikasi atribut yang memiliki kontribusi signifikan terhadap prediksi SLA. Seleksi fitur ini dilakukan menggunakan teknik ‘SelectFromModel’, dengan menggunakan ambang rata-rata

https://doi.org/10.35145/joisie.v9i1.4943

dari *importance score* sebagai batas seleksi. Setelah fitur-fitur penting ditentukan, data dibagi menjadi data latih dan data uji dengan rasio 80:20 menggunakan *stratified sampling* untuk menjaga proporsi kelas target. Mengingat adanya ketidakseimbangan kelas pada data target 'sla_met', maka dilakukan teknik *oversampling* dengan SMOTE (*Synthetic Minority Oversampling Technique*) pada data latih untuk menyeimbangkan jumlah data pada masing-masing kelas. Setelah itu, data dinormalisasi menggunakan 'StandardScaler' agar seluruh fitur memiliki skala nilai yang seragam dan tidak mendistorsi proses pelatihan.

2.4 *Data Mining*

Menurut Jessfry & Siddik, (2024) teknik *data mining* diharapkan mampu mempercepat proses pengambilan keputusan dalam perusahaan, sehingga dapat meningkatkan penjualan, meminimalkan kerugian, dan memperkuat daya saing terhadap perusahaan sejenis. Inti dari proses *Knowledge Discovery in Databases* (KDD) adalah *Data Mining*, di mana dilakukan eksplorasi data untuk menemukan pola dan membangun model prediktif (Jamalpur, 2020). Pada tahap ini, dilakukan pencarian parameter terbaik (*hyperparameter tuning*) untuk beberapa algoritma *machine learning* dengan menggunakan *library* Optuna. Optuna dipilih karena kemampuannya dalam melakukan optimasi secara efisien melalui pendekatan *Bayesian optimization* (Agrawal, 2021). Algoritma yang digunakan antara lain Logistic Regression, Random Forest, Decision Tree, dan XGBoost. Masing-masing algoritma dilatih menggunakan data yang telah diproses dan dievaluasi menggunakan berbagai metrik seperti akurasi, *precision, recall, F1-score*, dan AUC-ROC. Selain itu, *runtime* atau waktu komputasi juga dicatat sebagai bagian dari evaluasi efisiensi model. Proses ini menghasilkan model terbaik berdasarkan kombinasi kinerja prediktif dan efisiensi waktu.

2.5 Interpretasi dan Evaluasi

Interpretasi dan Evaluasi bertujuan untuk memahami bagaimana model membuat keputusan dan mengevaluasi fitur mana yang paling memengaruhi hasil prediksi. Salah satu teknik yang digunakan adalah SHAP (*SHapley Additive exPlanations*), yang memberikan penjelasan visual dan numerik terhadap kontribusi masing-masing fitur terhadap keputusan model. Dengan menggunakan SHAP, pengguna dapat mengetahui fitur apa saja yang paling berpengaruh terhadap keberhasilan SLA. Selain itu, juga dilakukan visualisasi *feature importance* berdasarkan model Decision Tree, yang memperlihatkan bobot masing-masing fitur dalam pengambilan keputusan.

3. HASIL DAN PEMBAHASAN

Tabel 1. Hasil Evaluasi Model Klasifikasi

Model	Accuracy	ROC AUC	Precision	Recall	F1 Score	Runtime (s)
XGBoost	0.9998	0.9854	0.9437	0.9710	0.9571	5.1439
Decision Tree	0.9818	0.9331	0.1095	0.8841	0.1949	0.8400
Random Forest	0.9957	0.9184	0.3515	0.8406	0.4957	26.1516
Logistic Regression	0.6910	0.6283	0.0045	0.5652	0.0090	0.1527

Dari hasil evaluasi pada Tabel 2. model XGBoost menunjukkan performa paling unggul dengan nilai akurasi sebesar 99.98%, ROC AUC 0.9854, serta F1 *score* sebesar 0.9571. Angka tersebut menunjukkan bahwa model ini tidak hanya mampu mengenali pola secara sangat akurat, namun juga memiliki kemampuan tinggi dalam membedakan kelas target positif dan negatif secara seimbang. Selain itu, nilai precision yang tinggi (0.9437) serta *recall* (0.9710) menunjukkan bahwa model ini sangat minim dalam memberikan

https://doi.org/10.35145/joisie.v9i1.4943

kesalahan klasifikasi, baik dalam bentuk *false positives* maupun *false negatives*. *Runtime*-nya pun tergolong efisien dibandingkan model ensemble lainnya.

Akurasi dan *precision* yang mencapai nilai 0.9998 pada model XGBoost dalam penelitian ini menunjukkan performa prediktif yang sangat tinggi. Meskipun angka ini tergolong mendekati sempurna, validitas hasil ini tidak dapat dianggap sebagai bentuk *overfitting* atau kesalahan pemodelan, mengingat proses pengolahan data, rekayasa fitur, serta tuning parameter telah dilakukan secara komprehensif dan sesuai dengan kaidah metodologis dalam *machine learning*. Pertama, seluruh data telah melalui tahapan pembersihan dan *preprocessing* yang sistematis, dimulai dari konversi format waktu ('resolved_at', 'closed_at') serta penghapusan entri yang tidak lengkap guna memastikan integritas temporal data. Selanjutnya, seluruh atribut kategorikal diencoding menggunakan teknik *Label Encoding*, memungkinkan transformasi yang seragam ke dalam bentuk numerik tanpa kehilangan struktur informasi. Selain itu, kolom-kolom yang tidak relevan seperti metadata dan identifier internal juga dieliminasi untuk menghindari terjadinya *data leakage*.

Dari sisi seleksi fitur, penelitian ini mengimplementasikan metode 'SelectFromModel' berbasis algoritma Random Forest untuk menyeleksi fitur-fitur penting secara otomatis. Seleksi ini bertujuan meningkatkan relevansi atribut terhadap target ('sla_met') serta menghindari kompleksitas model yang tidak perlu. Proses ini meningkatkan generalisasi dan ketahanan model terhadap noise. Selanjutnya, tantangan lain dalam klasifikasi biner yakni ketidakseimbangan kelas (*class imbalance*) ditangani melalui teknik *Synthetic Minority Over-sampling Technique* (SMOTE) yang menyeimbangkan proporsi data latih, sehingga model tidak mengalami bias terhadap kelas mayoritas. Proses pembagian data latih dan uji dilakukan secara *stratified sampling* dengan parameter 'random_state' yang dikunci (*seed* = 42) untuk menjaga konsistensi hasil dan reprodusibilitas eksperimen.

Model XGBoost dikembangkan melalui pendekatan *hyperparameter* tuning yang eksploratif menggunakan Optuna *Framework*, di mana fungsi objektif mengacu pada metrik ROC AUC untuk mencegah model hanya mengutamakan akurasi absolut. Hal ini memastikan model yang terbentuk tidak hanya memiliki performa tinggi secara angka, tetapi juga stabil dan tidak *overfit* terhadap data uji. Validitas akurasi dan *precision* yang tinggi semakin diperkuat oleh keseimbangan nilai metrik lain, seperti *recall* dan F1-*score*, yang menunjukkan bahwa model tidak hanya mampu mengenali kelas positif dengan sangat baik, tetapi juga meminimalkan *false positives* dan *false negatives.* Penambahan visualisasi berbasis SHAP (*SHapley Additive exPlanations*) juga menjadi kontribusi penting dalam memahami kontribusi masing-masing fitur terhadap keputusan model secara transparan dan interpretable. Dengan demikian, nilai akurasi sebesar 0.9998 yang dihasilkan bukan merupakan indikasi anomali, melainkan merupakan hasil dari serangkaian proses pemodelan yang robust, dan terstruktur. Implementasi metodologi ini menunjukkan bahwa model XGBoost yang dibangun dapat dijadikan referensi dalam sistem klasifikasi SLA yang bersifat kritis dan *real-time* dalam lingkungan operasional teknologi informasi

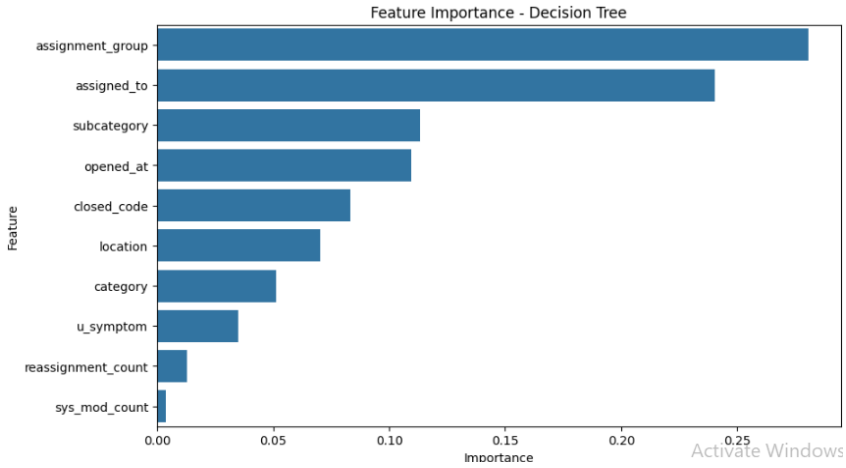
Model Decision Tree memiliki nilai akurasi sebesar 98.18%, dengan *recall* mencapai 88.41%, menunjukkan bahwa model ini sensitif terhadap kasus-kasus positif. Namun demikian, *precision*-nya sangat rendah, yaitu hanya 10.95%, yang berarti banyak prediksi positif dari model yang ternyata salah. F1-*score*-nya pun hanya 0.1949, yang menunjukkan ketidakseimbangan antara *precision* dan *recall*. Meskipun begitu, model ini memiliki kelebihan dalam *runtime* yang cepat, yaitu hanya 0.84 detik, menjadikannya pilihan yang tepat ketika dibutuhkan kecepatan dengan interpretasi model yang mudah.

Model Random Forest memberikan hasil yang cukup baik, dengan akurasi sebesar 99.57%, namun *precision* dan *recall* berada pada tingkat menengah, yaitu masing-masing 35.15% dan 84.06%. F1 *score*-nya sebesar 0.4957 menunjukkan bahwa model ini lebih baik daripada Decision Tree dalam keseimbangan klasifikasi, namun belum optimal seperti XGBoost. Salah satu kelemahan model ini adalah waktu proses yang relatif lama, mencapai lebih dari 26 detik, karena banyaknya pohon keputusan yang dibentuk dalam proses ensemble.

Model Logistic Regression menunjukkan performa terendah dari semua model yang diuji. Akurasinya hanya 69.10%, dengan precision sangat rendah, yaitu 0.0045, dan F1 score hanya 0.0090. Meskipun recall mencapai 56.52%, angka tersebut tidak dapat menutupi kelemahan besar model ini dalam mendeteksi dengan tepat kasus positif. Ini menandakan bahwa Logistic Regression tidak cocok digunakan untuk dataset dengan karakteristik kompleks atau distribusi yang tidak seimbang.

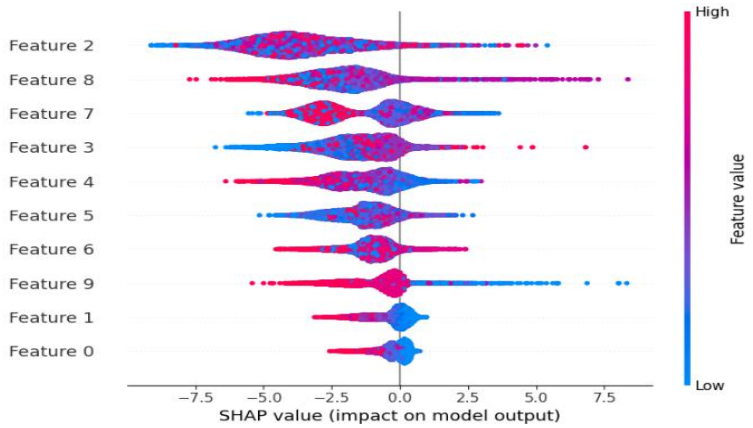
https://doi.org/10.35145/joisie.v9i1.4943

Untuk mendukung interpretabilitas model, dilakukan analisis terhadap kontribusi masing-masing fitur menggunakan feature importance dan SHAP *analysis* seperti pada gambar 2. Pada model Decision Tree, fitur-fitur yang paling dominan adalah 'assignment_group', 'assigned_to', 'subcategory', dan 'opened_at'. Fitur 'assignment_group' dan 'assigned_to' merupakan variabel yang menunjukkan penanggung jawab terhadap tiket insiden, yang secara logis sangat berkorelasi terhadap kecepatan dan keberhasilan penyelesaian kasus. Hal ini mengindikasikan bahwa struktur organisasi dan alokasi tugas memiliki peranan penting dalam hasil akhir insiden.



Gambar 2. Grafik fitur yang paling dominan

Pada model XGBoost, dilakukan analisis menggunakan SHAP summary plot seperti pada gambar 3. Visualisasi ini memberikan wawasan yang lebih dalam mengenai seberapa besar dan ke arah mana suatu fitur mempengaruhi output prediksi model. Fitur dengan nilai SHAP tertinggi menunjukkan bahwa perubahan nilainya secara signifikan mempengaruhi hasil prediksi. Dalam visualisasi SHAP, warna merah menandakan nilai fitur tinggi, sedangkan biru menandakan nilai rendah. Hasil ini menunjukkan bahwa XGBoost tidak hanya unggul dalam akurasi, tetapi juga sangat kuat dari sisi interpretasi dan transparansi model, yang penting dalam konteks sistem pengambilan keputusan berbasis AI.



Gambar 3. SHAP *Summary Plot*

4. KESIMPULAN

Penelitian ini menunjukkan bahwa algoritma XGBoost merupakan model terbaik dalam melakukan klasifikasi terhadap kepatuhan *Service Level Agreement* (SLA) berdasarkan data *incident log*. Dengan akurasi mencapai 99,98%, F1-score sebesar 0,9571, *precision* 0,9437, dan *recall* 0,9710, XGBoost menunjukkan

https://doi.org/10.35145/joisie.v9i1.4943

performa yang sangat unggul dibandingkan dengan model lainnya, seperti Random Forest, Decision Tree, dan Logistic Regression. Hasil evaluasi ini mengindikasikan bahwa XGBoost tidak hanya mampu mengenali pola secara akurat, tetapi juga menunjukkan kemampuan yang seimbang dalam membedakan kelas positif dan negatif, serta memiliki sensitivitas tinggi terhadap kelas minoritas (SLA tidak terpenuhi). Dengan kemampuan klasifikasi yang tinggi dan interpretasi yang transparan, model ini berpotensi menjadi alat bantu strategis bagi manajer layanan TI dalam mengidentifikasi insiden yang berisiko tinggi gagal memenuhi SLA, sehingga tindakan korektif dapat dilakukan secara proaktif untuk meningkatkan kepuasan pengguna dan kualitas layanan.

Kinerja XGBoost tidak diperoleh secara instan, melainkan merupakan hasil dari proses pemodelan yang ketat, mulai dari tahap pembersihan data, *encoding* fitur kategorikal, seleksi fitur berbasis algoritma Random Forest, hingga penanganan ketidakseimbangan data menggunakan SMOTE. *Hyperparameter tuning* melalui Optuna dengan acuan ROC AUC turut memperkuat kemampuan generalisasi model. Selain itu, penggunaan visualisasi SHAP memberikan nilai tambah dari segi interpretabilitas, menjadikan XGBoost tidak hanya presisi secara numerik, tetapi juga transparan dalam proses pengambilan keputusan.

Model lain seperti Decision Tree dan Random Forest menunjukkan performa sedang, dengan akurasi masing-masing sebesar 98,18% dan 99,57%. Namun, nilai *precision* dan F1-*score* yang lebih rendah menjadikannya kurang ideal dalam konteks klasifikasi SLA yang menuntut sensitivitas tinggi terhadap kasus ketidakpatuhan. Sementara itu, Logistic Regression terbukti tidak cocok diterapkan pada dataset ini, karena menghasilkan akurasi dan precision yang sangat rendah serta ketidakseimbangan dalam klasifikasi positif dan negatif.

Penelitian lanjutan disarankan untuk memperluas cakupan data dan mempertimbangkan integrasi fitur kontekstual tambahan guna meningkatkan ketepatan prediksi. Salah satu arah potensial adalah pemanfaatan data deskripsi insiden dalam bentuk teks menggunakan teknik *Natural Language Processing* (NLP), serta memasukkan data historis terkait kinerja agen penyelesaian tiket sebagai fitur prediktif. Selain itu, meskipun model XGBoost telah menunjukkan performa unggul, eksplorasi lebih lanjut terhadap model *ensemble* modern lainnya seperti LightGBM dan CatBoost, maupun pendekatan *ensemble hybrid* seperti *stacking*, dapat membuka peluang peningkatan akurasi dan stabilitas model. Evaluasi performa model secara jangka panjang juga penting untuk menjamin keberlanjutan kualitas prediksi terhadap data baru. Penerapan sistem pemantauan model (*model drift monitoring*) sangat disarankan agar deteksi perubahan pola data dapat dilakukan secara dini dan model dapat segera disesuaikan dengan kondisi operasional terkini.

5. DAFTAR PUSTAKA

Agrawal, T. (2021). *Optuna and AutoML BT - Hyperparameter Optimization in Machine Learning: Make Your Machine Learning and Deep Learning Models More Efficient* (T. Agrawal (ed.); pp. 109–129). Apress. https://doi.org/10.1007/978-1-4842-6579-6_5

Amaral, C., Fantinato, M., & Peres, S. (2018). Incident management process enriched event log. In *UCI Machine Learning Repository*. https://doi.org/10.24432/C57S4H

Amazonas, J. R., Akbari-Moghanjoughi, A., Santos-Boada, G., & Solé Pareta, J. (2019). Service Level Agreements for Communication Networks: A Survey. *INFOCOMP Journal of Computer Science*, *18*(1 SE-), 32–56. https://infocomp.dcc.ufla.br/index.php/infocomp/article/view/584

Ansori, M. A., Anshori, M. F., Pratama, A. W., & Gunawan, A. (2023). Prediksi Waktu Tempuh SLA (Service Level Agreement) Pengiriman PT. Pos Indonesia Menggunakan Random Forest. *Jurnal Sistem Informasi*, *6*(1), 39–46. https://doi.org/10.34128/jsi.v6i1.209

Bhargavi, N., & Sekhar, R. R. (2021). An architectural approach to provide efficient service with the Service Level Agreement. *International Journal of Computer Science, Engineering, and Information Technology*, *7*(1), 1–8. https://www.academia.edu/download/65870068/CSEIT217133.pdf

Dachi, J. M. A. S., & Sitompul, P. (2023). Analisis Perbandingan Algoritma XGBoost dan Algoritma Random Forest Ensemble Learning pada Klasifikasi Keputusan Kredit. *Jurnal Riset Rumpun Matematika Dan Ilmu Pengetahuan Alam*, *2*(2), 87–103. https://doi.org/10.55606/jurrimipa.v2i2.1470

Dewi, S. P., Nurwati, N., & Rahayu, E. (2022). Penerapan Data Mining Untuk Prediksi Penjualan Produk Terlaris Menggunakan Metode K-Nearest Neighbor. *Building of Informatics, Technology and Science (BITS)*, *3*(4), 639–648. https://doi.org/10.47065/bits.v3i4.1408

https://doi.org/10.35145/joisie.v9i1.4943

Didavi, A. B. K., Agbokpanzo, R. G., & Agbomahena, M. (2021). Comparative study of Decision Tree, Random Forest and XGBoost performance in forecasting the power output of a photovoltaic system. *Proceedings of the 4th International Conference on Bio-Engineering for Smart Technologies (BioSMART)*, 1–5. https://doi.org/10.1109/BioSMART54244.2021.9677566

Ergün, S. (2023). Explaining xgboost predictions with shap value: A comprehensive guide to interpreting decision tree-based models. *New Trends in Computer Sciences*, *1*(1), 19–31. https://doi.org/10.3846/ntcs.2023.17901

Gonçalves, J. P. de B., Gomes, R. L., Villaca, R. da S., Municio, E., & Marquez-Barja, J. (2020). A Service Level Agreement Verification System using Blockchains. *2020 IEEE 11th International Conference on Software Engineering and Service Science (ICSESS)*, 541–544. https://doi.org/10.1109/ICSESS49938.2020.9237735

Hou, C., Cao, B., & Fan, J. (2022). A data-driven method to predict service level for call centers. *IET Communications*, *16*(10), 1241–1252. https://doi.org/https://doi.org/10.1049/cmu2.12192

Jamalpur, B. (2020). Data Exploration As a Process of Knowledge Finding and the Role of Mining Data Towards Information Security. *Journal of Mechanics of Continua and Mathematical Sciences*, *15*(6), 213–223. https://doi.org/10.26782/jmcms.2020.06.00017

Jessfry, V., & Siddik, M. (2024). Penerapan Data Mining Menggunakan Algortima Apriori Dalam Membangun Sistem Persediaan Barang. *Journal Of Information Systems And Informatics Engineering*, *8*(1), 187–199.

Meneses, C. J., & Grinstein, G. G. (1998). *Transactions on Information and Communications Technologies vol 19 © 1998 WIT Press, www.witpress.com, ISSN 1743-3517*. *19*, 1–7.

Nugraha, Y., Martin, A., Nugraha, Y., Martin, A., Trustworthy, U., Level, S., & Open, A. (2018). *Understanding Trustworthy Service Level Agreements : Open Problems and Existing Solutions To cite this version : HAL Id : hal-01684231 Understanding Trustworthy Service Level Agreements : Open Problems and Existing Solutions*.

Nur, N., Wajidi, F., Sulfayanti, S., & Wildayani, W. (2023). Implementasi Algoritma Random Forest Regression untuk Memprediksi Hasil Panen Padi di Desa Minanga. *Jurnal Komputer Terapan*, *9*(1 SE-), 58–64. https://doi.org/10.35143/jkt.v9i1.5917

Sahai, R., Al-Ataby, A., Assi, S., Jayabalan, M., Liatsis, P., Loy, C. K., Al-Hamid, A., Al-Sudani, S., Alamran, M., & Kolivand, H. (2023). *Insurance Risk Prediction Using Machine Learning BT - Data Science and Emerging Technologies* (Y. B. Wah, M. W. Berry, A. Mohamed, & D. Al-Jumeily (eds.); pp. 419–433). Springer Nature Singapore.

Suja, T. L., & Booba, B. (2020). *A feasibility study of service level agreement compliance for start-ups. In Data Management, Analytics and Innovation*. Proceedings of ICDMAI 2020. https://books.google.com/books?hl=en&lr=&id=lFj4DwAAQBAJ&oi=fnd&pg=PA407

Supono, S. (2020). Model Penilaian Kapabilitas Proses Layanan Service Level Agreement (SLA) Pada Cloud Computing. *Jurnal Sains Dan Informatika*, *6*(1), 62–71. https://doi.org/10.34128/jsi.v6i1.209

Tai, T.-E., Haw, S.-C., Ng, K.-W., Naveen, P., & Al-Tarawneh, M. (2023). Performance Evaluation on Resolution Time Prediction Using Decision Tree, Random Forest and eXtreme Gradient Boosting. *2023 International Conference on Computer Applications Technology (CCAT)*, 74–79. https://doi.org/10.1109/CCAT59108.2023.00021

Tan, W., Hai, Z., Jinjing, T., Yao, Z., Li Da, X., & and Guo, K. (2022). A novel service level agreement model using blockchain and smart contract for cloud manufacturing in industry 4.0. *Enterprise Information Systems*, *16*(12), 1939426. https://doi.org/10.1080/17517575.2021.1939426

Yudistira, N., & Putra, A. F. (2021). Algoritma Decision Tree Dan Smote Untuk Klasifikasi Serangan Jantung Miokarditis Yang Imbalance. *Jurnal Litbang Edusaintech*, *2*(2), 112–122. https://doi.org/10.51402/jle.v2i2.48