

PENERAPAN ALGORITMA *ISOLATION FOREST* UNTUK DETEKSI ANOMALI DATA PELAYANAN PERIZINAN DI DPMPTSP KABUPATEN ACEH TENGGARA

Khozi Shaid Almunawar¹, Hevlie Winda Nazry Simbolon²

^{1,2}Universitas Muhammadiyah Sumatera Utara, Indonesia

Email: khozishaid@gmail.com

ABSTRAK

Penelitian ini bertujuan untuk menerapkan algoritma *Isolation Forest* dalam mendeteksi anomali pada data pelayanan perizinan di Dinas Penanaman Modal dan Pelayanan Terpadu Satu Pintu (DPMPTSP) Kabupaten Aceh Tenggara serta mengembangkan sistem deteksi anomali berbasis web. Latar belakang penelitian ini didasarkan pada meningkatnya volume data pelayanan perizinan yang dapat menimbulkan permasalahan dalam proses pengelolaan dan pemantauan data, seperti keterlambatan proses pelayanan, ketidakkonsistenan data, kesalahan input, dan data yang memiliki karakteristik berbeda dari pola umum. Proses identifikasi anomali secara manual pada data dalam jumlah besar membutuhkan waktu dan ketelitian sehingga diperlukan metode yang mampu melakukan deteksi secara otomatis dan sistematis.

Metode penelitian yang digunakan adalah metode kuantitatif dengan pendekatan eksperimen komputasional berbasis machine learning. Data yang digunakan merupakan data sekunder berupa data pelayanan perizinan DPMPTSP Kabupaten Aceh Tenggara periode tahun 2024 sebanyak 1.000 data. Tahapan penelitian meliputi pengumpulan data, data preprocessing, data cleaning, pemeriksaan missing value, pemeriksaan data duplikat, perhitungan Lama Proses, dan transformasi data kategorikal menggunakan Label Encoding. Selanjutnya, data dianalisis menggunakan algoritma *Isolation Forest* yang bekerja dengan mengisolasi data melalui proses partisi secara acak pada sejumlah *Isolation Tree*. Data yang lebih mudah terisolasi dari mayoritas data memiliki karakteristik yang lebih menyimpang dan memperoleh tingkat penyimpangan yang lebih tinggi. Hasil deteksi kemudian dianalisis berdasarkan jumlah data normal dan anomali, anomaly score, serta karakteristik data yang teridentifikasi sebagai anomali.

Sistem yang dikembangkan berupa aplikasi berbasis web menggunakan bahasa pemrograman Python dan framework Flask. Sistem dirancang untuk memungkinkan pengguna mengunggah data pelayanan perizinan, melakukan proses deteksi anomali, serta menampilkan hasil analisis dalam bentuk tabel, label data normal dan anomali, nilai *anomaly score*, dan visualisasi hasil deteksi. Melalui penerapan algoritma *Isolation Forest*, penelitian ini diharapkan dapat membantu proses identifikasi data pelayanan yang memiliki karakteristik menyimpang secara lebih cepat, sistematis, dan objektif serta dapat menjadi pendukung dalam evaluasi kualitas pengelolaan data pelayanan perizinan di DPMPTSP Kabupaten Aceh Tenggara.

Kata kunci: *Isolation Forest*, Deteksi Anomali, Data Pelayanan Perizinan, *Machine Learning*, DPMPTSP.

ABSTRACT

This study aims to apply the Isolation Forest algorithm to detect anomalies in licensing service data at the Investment and One-Stop Integrated Services Office (DPMPTSP) of Southeast Aceh Regency and to develop a web-based anomaly detection system. The background of this study is based on the increasing volume of licensing

service data, which can lead to various problems in data management and monitoring, such as delays in service processes, data inconsistencies, input errors, and data with characteristics that differ from general patterns. Manually identifying anomalies in large amounts of data requires considerable time and accuracy; therefore, a method capable of detecting anomalies automatically and systematically is needed.

The research method used in this study is a quantitative method with a computational experimental approach based on machine learning. The data used are secondary data consisting of licensing service data from DPMPTSP Southeast Aceh Regency for the 2024 period, totaling 1,000 records. The research stages include data collection, data preprocessing, data cleaning, missing value checking, duplicate data checking, calculation of service processing time, and transformation of categorical data using Label Encoding. The data were then analyzed using the Isolation Forest algorithm, which works by isolating data through random partitioning across a number of Isolation Trees. Data that are more easily isolated from the majority tend to have more distinctive characteristics and a higher degree of deviation. The detection results were then analyzed based on the number of normal and anomalous data, anomaly scores, and the characteristics of the identified anomalies.

The developed system is a web-based application using the Python programming language and the Flask framework. The system is designed to allow users to upload licensing service data, perform the anomaly detection process, and display the analysis results in the form of tables, normal and anomaly labels, anomaly scores, and visualizations of the detection results. Through the application of the Isolation Forest algorithm, this study is expected to assist in identifying licensing service data with abnormal characteristics more quickly, systematically, and objectively, as well as support the evaluation of data management quality in licensing services at DPMPTSP Southeast Aceh Regency.

Keywords: Isolation Forest, Anomaly Detection, Licensing Service Data, Machine Learning, DPMPTSP.

PENDAHULUAN

Perkembangan teknologi informasi di era digital mendorong instansi pemerintah untuk meningkatkan kualitas pelayanan publik melalui pemanfaatan sistem elektronik. Salah satu instansi yang memiliki peran strategis dalam penyelenggaraan pelayanan perizinan adalah Dinas Penanaman Modal dan Pelayanan Terpadu Satu Pintu (DPMPTSP), yang menangani berbagai layanan, seperti Persetujuan Bangunan Gedung (PBG), Surat Izin Praktik (SIP), Surat Izin Usaha Perdagangan (SIUP), Surat Izin Usaha Jasa Konstruksi (SIUJK), dan Izin Usaha Industri (IUI). Peningkatan kebutuhan masyarakat terhadap layanan perizinan berdampak pada bertambahnya volume data yang harus dikelola. Kondisi tersebut menimbulkan tantangan dalam menjaga kualitas data, mengidentifikasi penyimpangan, dan meningkatkan efektivitas pelayanan. Permasalahan ini semakin kompleks karena akumulasi data berpotensi memuat keterlambatan penyelesaian layanan, inkonsistensi informasi, kesalahan penginputan, dan duplikasi data. Pemeriksaan secara manual juga menjadi kurang efektif ketika jumlah data terus meningkat karena membutuhkan waktu dan ketelitian yang tinggi.

Berdasarkan observasi di DPMPTSP Kabupaten Aceh Tenggara, penyelenggaraan pelayanan perizinan telah didukung oleh sistem elektronik yang memungkinkan penyimpanan data permohonan secara berkelanjutan. Namun, peningkatan volume data menimbulkan kendala dalam proses pemantauan, terutama untuk mengidentifikasi data dengan karakteristik tidak lazim, seperti durasi penyelesaian perizinan yang melampaui waktu normal, ketidaksesuaian informasi,

dan kemungkinan kesalahan penginputan. Pemeriksaan terhadap kondisi tersebut masih memerlukan proses manual yang relatif memakan waktu. Dalam konteks ini, anomali dapat dipahami sebagai data atau kondisi yang menyimpang dari pola mayoritas, misalnya durasi pelayanan yang jauh lebih panjang dibandingkan dengan kecenderungan umum. Anomali tidak secara langsung menunjukkan adanya kesalahan atau pelanggaran, tetapi dapat menjadi indikasi awal yang memerlukan verifikasi lebih lanjut. Oleh karena itu, diperlukan pendekatan analisis yang mampu mengidentifikasi penyimpangan secara otomatis, cepat, dan terstruktur untuk mendukung evaluasi kualitas pengelolaan data pelayanan perizinan.

Salah satu pendekatan yang dapat digunakan untuk mengatasi permasalahan tersebut adalah *machine learning*, khususnya teknik *anomaly detection*, yaitu metode untuk menemukan data yang memiliki karakteristik berbeda dari pola mayoritas dalam suatu kumpulan data. Alang & Bindewari (2025) menunjukkan bahwa pemanfaatan algoritma *machine learning* dapat meningkatkan kemampuan identifikasi anomali, terutama pada data berukuran besar dan memiliki kompleksitas tinggi. Dalam pelayanan publik, deteksi anomali dapat membantu mengenali durasi pemrosesan perizinan yang berlebihan, ketidakkonsistenan informasi, dan pola pelayanan yang tidak lazim. Informasi yang dihasilkan dapat digunakan untuk menentukan data yang perlu diverifikasi lebih lanjut sekaligus mendukung pengambilan keputusan berbasis data. Dengan demikian, penerapan *anomaly detection* berpotensi membantu instansi melakukan pemantauan dan evaluasi pelayanan secara lebih terarah.

Salah satu algoritma yang dapat diterapkan adalah *Isolation Forest*, yang mengidentifikasi anomali melalui pemisahan data secara acak menggunakan struktur pohon yang disebut *Isolation Tree*. Data yang memiliki karakteristik berbeda dari mayoritas cenderung dapat diisolasi melalui jumlah tahapan yang lebih sedikit, sehingga menghasilkan nilai *anomaly score* untuk menggambarkan tingkat indikasi penyimpangannya. Menurut Liu et al. (2024), *Isolation Forest* memiliki efisiensi komputasi yang baik, mampu menangani data berdimensi besar, dan tidak mensyaratkan label kelas dalam proses pelatihan karena menggunakan pendekatan *unsupervised learning*. Karakteristik tersebut relevan dengan data pelayanan perizinan yang belum memiliki klasifikasi awal antara data normal dan anomali. Selain itu, Zhang & Liu (2022) menjelaskan bahwa kecepatan pemrosesan dan efisiensi *Isolation Forest* dalam menangani data berukuran besar mendukung pemanfaatannya dalam berbagai penelitian deteksi anomali modern. Dengan demikian, algoritma ini dapat menjadi alternatif untuk mengidentifikasi penyimpangan berdasarkan pola yang terdapat dalam data tanpa memerlukan pelabelan awal.

Berdasarkan permasalahan pengelolaan data pelayanan perizinan dan karakteristik algoritma tersebut, penelitian ini menerapkan *Isolation Forest* untuk mendeteksi anomali pada data pelayanan perizinan DPMPTSP Kabupaten Aceh Tenggara. Penerapan algoritma ini diarahkan untuk mengidentifikasi data yang memiliki karakteristik berbeda dari pola mayoritas, menganalisis distribusi anomali

berdasarkan atribut pelayanan, serta menyajikan informasi yang dapat digunakan sebagai dasar pemeriksaan lebih lanjut. Hasil deteksi tidak dimaksudkan sebagai bukti langsung adanya kesalahan pelayanan, melainkan sebagai indikasi awal untuk mendukung verifikasi dan evaluasi kualitas data. Melalui pendekatan *machine learning*, penelitian ini diharapkan dapat mendukung pemantauan data secara lebih sistematis dan membantu instansi memperoleh informasi yang relevan untuk meningkatkan efektivitas pengelolaan pelayanan perizinan.

METODE PENELITIAN

Penelitian ini menggunakan metode kuantitatif dengan pendekatan eksperimen komputasional berbasis *machine learning* untuk mengidentifikasi data anomali pelayanan perizinan menggunakan algoritma *Isolation Forest*. Penelitian dilaksanakan di Dinas Penanaman Modal dan Pelayanan Terpadu Satu Pintu (DPMPTSP) Kabupaten Aceh Tenggara pada Februari hingga Juni 2026. Data diperoleh melalui dokumentasi dataset pelayanan perizinan dan studi literatur yang relevan. Dari sembilan atribut awal, Id Pemohon tidak digunakan dalam pemodelan, sedangkan Tgl. Terbit OSS dan Tgl. Izin digunakan untuk menghitung variabel Lama Proses. Fitur yang digunakan dalam pemodelan meliputi Lama Proses, Kecamatan, KBLI, Kode Risiko, Jenis Perizinan, Status Perizinan, dan KL/Sektor. Perancangan aplikasi berbasis web dilakukan menggunakan *Unified Modeling Language* (UML), yang mencakup *Use Case Diagram*, *Activity Diagram*, *Sequence Diagram*, dan *Class Diagram*.

Tahapan penelitian meliputi pengumpulan data, prapemrosesan, penerapan algoritma *Isolation Forest*, deteksi anomali, analisis hasil, visualisasi, dan implementasi aplikasi berbasis web. Prapemrosesan dilakukan melalui pemeriksaan data duplikat, data tidak valid, ketidakkonsistenan format, dan nilai kosong. Lama Proses dihitung berdasarkan selisih Tgl. Izin dan Tgl. Terbit OSS. Dari 1.000 data awal, satu data dengan durasi -1 hari dikeluarkan karena menunjukkan ketidaksesuaian urutan tanggal, sehingga tersisa 999 data untuk pemodelan. Atribut kategorikal ditransformasikan menggunakan teknik *Label Encoding*, sedangkan KBLI yang telah berbentuk numerik tidak memerlukan pengkodean. Algoritma *Isolation Forest* diterapkan dengan parameter *n_estimators* sebesar 100, *max_samples* bernilai *auto*, *contamination* sebesar 0,05, dan *random_state* sebesar 42. Model menghasilkan *anomaly score* dan label prediksi, dengan label -1 menunjukkan data anomali dan label 1 menunjukkan data normal.

Analisis dilakukan menggunakan statistik deskriptif, pemeriksaan *anomaly score*, identifikasi karakteristik anomali, dan visualisasi hasil deteksi. Analisis mencakup jumlah dan proporsi data normal serta anomali, distribusi durasi pelayanan, dan persebaran anomali berdasarkan kecamatan, KBLI, serta karakteristik perizinan lainnya. Hasil analisis disajikan melalui grafik perbandingan data normal dan anomali, diagram proporsi, distribusi *anomaly score*, *boxplot* Lama Proses, grafik kecamatan dan KBLI dengan jumlah anomali terbanyak, *scatter plot* hubungan Lama

Proses dengan *anomaly score*, serta *timeline* kemunculan anomali berdasarkan bulan. Data yang terdeteksi sebagai anomali diperlakukan sebagai indikasi penyimpangan dari pola mayoritas, bukan sebagai bukti langsung adanya kesalahan pelayanan. Seluruh hasil digunakan untuk mendukung evaluasi data dan penyajian informasi pada aplikasi deteksi anomali berbasis web di DPMPTSP Kabupaten Aceh Tenggara.

HASIL DAN PEMBAHASAN

Karakteristik Dataset dan Hasil Preprocessing

Dataset yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari Dinas Penanaman Modal dan Pelayanan Terpadu Satu Pintu (DPMPTSP) Kabupaten Aceh Tenggara. Data tersebut berasal dari sistem *Online Single Submission* (OSS) yang berisi informasi mengenai pelayanan perizinan yang telah diproses oleh DPMPTSP. Dataset digunakan sebagai sumber data utama dalam penerapan algoritma *Isolation Forest* untuk mendeteksi data yang memiliki karakteristik berbeda dari sebagian besar data lainnya atau anomali.

Data yang digunakan merupakan data pelayanan perizinan DPMPTSP Kabupaten Aceh Tenggara periode Januari 2024 hingga September 2024. Dataset awal terdiri atas 1.000 data yang memuat informasi mengenai identitas permohonan, tanggal penerbitan OSS, tanggal penerbitan izin, lokasi usaha, klasifikasi usaha, tingkat risiko usaha, jenis perizinan, status perizinan, serta kementerian atau lembaga sektor yang berwenang.

Tahap *preprocessing* dilakukan untuk mempersiapkan dataset sebelum digunakan dalam proses pembentukan model *Isolation Forest*. Tahapan ini meliputi *data cleaning*, perhitungan lama proses pelayanan, dan transformasi data kategorikal menggunakan teknik *Label Encoding*. Hasil dari tahapan *preprocessing* digunakan sebagai dasar dalam proses pembentukan model deteksi anomali.

Hasil pemeriksaan terhadap 1.000 data awal menunjukkan bahwa tidak terdapat missing value dan tidak terdapat data duplikat. Selanjutnya dilakukan perhitungan Lama Proses berdasarkan selisih antara Tanggal Izin dan Tanggal Terbit OSS. Satu data dengan nilai Lama Proses -1 hari dinyatakan sebagai data tidak valid dan dikeluarkan dari dataset sebelum proses pembentukan model. Setelah data tersebut dikeluarkan, jumlah data yang digunakan untuk proses pemodelan menjadi 999 data.

Dengan demikian, proses data cleaning menghasilkan perubahan jumlah data dari 1.000 data awal menjadi 999 data yang memenuhi kriteria untuk digunakan dalam proses pemodelan. Setelah pemeriksaan *missing value*, dilakukan pemeriksaan data duplikat. Pemeriksaan ini bertujuan untuk mengetahui apakah terdapat data yang tercatat lebih dari satu kali sehingga dapat memengaruhi hasil analisis.

Hasil pemeriksaan menunjukkan bahwa tidak ditemukan data duplikat pada dataset penelitian. Dengan demikian, tidak diperlukan proses penghapusan data berdasarkan duplikasi.

Selanjutnya dilakukan pemeriksaan tipe data untuk memastikan bahwa setiap

atribut memiliki tipe data yang sesuai dengan proses pengolahan. Atribut tanggal digunakan dalam perhitungan Lama Proses, atribut numerik digunakan sebagai fitur numerik, sedangkan atribut kategorikal ditransformasikan menjadi bentuk numerik menggunakan *Label Encoding*. Hasil pemeriksaan menunjukkan bahwa atribut tanggal telah memiliki tipe data tanggal, sedangkan atribut kategorikal dan numerik dapat diproses sesuai kebutuhan penelitian.

Pembentukan Model *Isolation Forest*

Model dibangun menggunakan implementasi *Isolation Forest* pada pustaka Scikit-learn di lingkungan Python. Konfigurasi model mencakup 100 pohon isolasi ($n_estimators = 100$), $max_samples = auto$, $contamination = 0,05$, dan $random_state = 42$. *Isolation Forest* mengidentifikasi observasi yang relatif mudah diisolasi melalui pohon-pohon acak sebagai kandidat anomali. Parameter $contamination$ ditetapkan sebesar 5%, sehingga model dikonfigurasi untuk menandai proporsi observasi yang mendekati nilai tersebut sebagai anomali. Parameter $random_state$ digunakan untuk mendukung keterulangan hasil.

Tabel 1. Parameter Model *Isolation Forest*

Parameter	Nilai	Fungsi
$n_estimators$	100	Jumlah pohon isolasi yang dibangun.
$max_samples$	auto	Menentukan jumlah sampel yang digunakan pada setiap pohon secara otomatis.
$contamination$	0,05	Perkiraan proporsi anomali yang digunakan untuk menentukan ambang deteksi.
$random_state$	42	Mendukung konsistensi hasil ketika proses dijalankan kembali.

Penggunaan $contamination$ sebesar 0,05 merupakan keputusan konfigurasi, bukan bukti bahwa proporsi anomali sebenarnya pada populasi pelayanan memang 5%. Karena penelitian tidak memiliki label acuan (*ground truth*), hasil model perlu dipahami sebagai penandaan awal berdasarkan pola dalam dataset, bukan sebagai klasifikasi yang telah terverifikasi.

Hasil Deteksi dan Distribusi Anomaly Score

Penerapan model terhadap 999 observasi menghasilkan 949 data berlabel normal (1) dan 50 data berlabel anomali (-1). Proporsi tersebut setara dengan 94,99% data normal dan 5,01% data anomali. Jumlah ini mendekati proporsi yang ditetapkan melalui parameter $contamination$. Persentase tersebut menggambarkan distribusi label keluaran model dan tidak dapat diperlakukan sebagai nilai akurasi.

Tabel 2. Hasil Deteksi Anomali

Kategori hasil model	Jumlah	Persentase
Normal (label 1)	949	94,99%
Anomali (label -1)	50	5,01%
Total	999	100%

Nilai anomaly score berada pada rentang $-0,164676$ hingga $0,283513$, dengan rata-rata $0,207643$ dan simpangan baku $0,094145$. Kuartil pertama tercatat sebesar $0,173701$, median $0,247079$, dan kuartil ketiga $0,269560$. Dalam keluaran yang dilaporkan pada penelitian ini, skor yang lebih rendah menunjukkan observasi yang lebih menyimpang dari pola mayoritas, sedangkan skor yang lebih tinggi menunjukkan karakteristik yang relatif lebih dekat dengan pola mayoritas. Label prediksi tetap digunakan untuk menentukan kategori normal atau anomali, sementara skor berfungsi sebagai informasi pendukung untuk menelaah tingkat penyimpangan.

Tabel 3. Skor Statistik Anomaly

Statistik anomaly score	Nilai
Jumlah observasi	999
Rata-rata	$0,207643$
Simpangan baku	$0,094145$
Minimum	$-0,164676$
Kuartil 1 (25%)	$0,173701$
Median (50%)	$0,247079$
Kuartil 3 (75%)	$0,269560$
Maksimum	$0,283513$

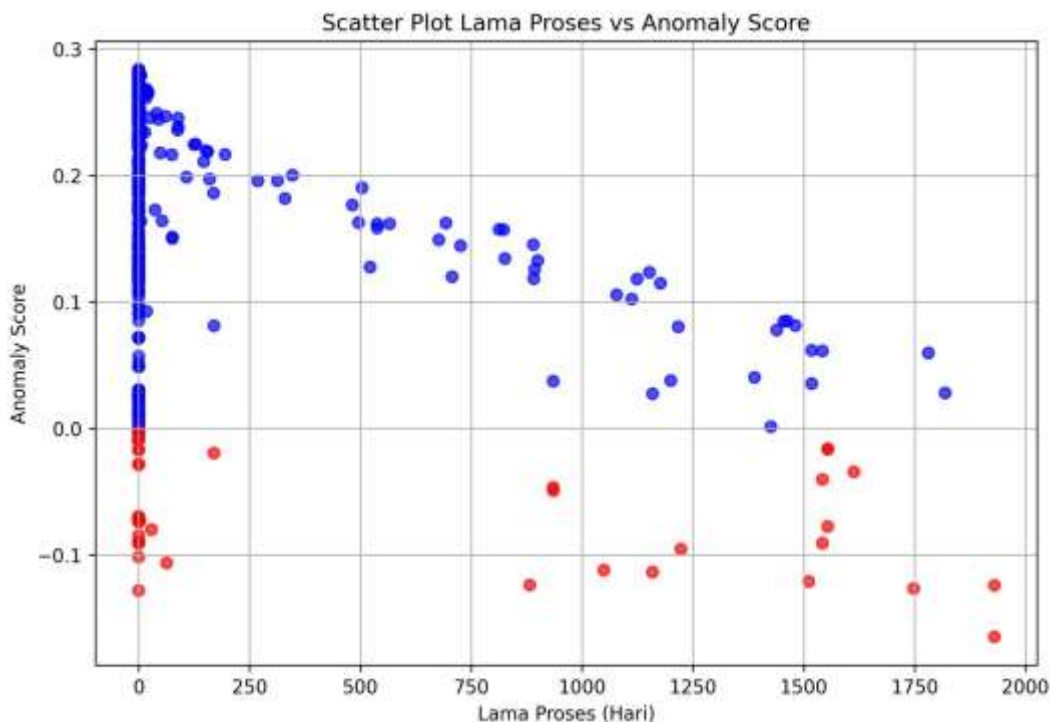
Karakteristik Data yang Teridentifikasi sebagai Anomali

Analisis terhadap 50 observasi yang ditandai sebagai anomali menunjukkan bahwa sebagian di antaranya memiliki durasi pelayanan yang jauh lebih panjang daripada mayoritas data. Durasi yang dilaporkan antara lain 882, 935, 1.049, 1.223, 1.511, 1.542, 1.554, 1.613, 1.748, dan hingga 1.930 hari. Rentang tersebut memperlihatkan variasi durasi yang besar dan layak menjadi prioritas pemeriksaan administratif. Meskipun demikian, durasi yang panjang tidak menjadi satu-satunya penentu anomali karena model mempertimbangkan keseluruhan fitur yang dimasukkan.

Distribusi anomali berdasarkan kecamatan memperlihatkan jumlah tertinggi pada Kecamatan Babussalam (11 data), diikuti Lawe Bulan (7), Semadam (6), dan Babel (5). Sebaran ini menunjukkan bahwa observasi yang berbeda dari pola mayoritas tidak terdistribusi secara merata pada wilayah-wilayah tersebut. Namun, angka tersebut merupakan jumlah absolut, bukan tingkat anomali relatif terhadap seluruh permohonan pada setiap kecamatan. Tanpa membandingkan jumlah anomali dengan total permohonan per kecamatan, tidak tepat menyimpulkan bahwa

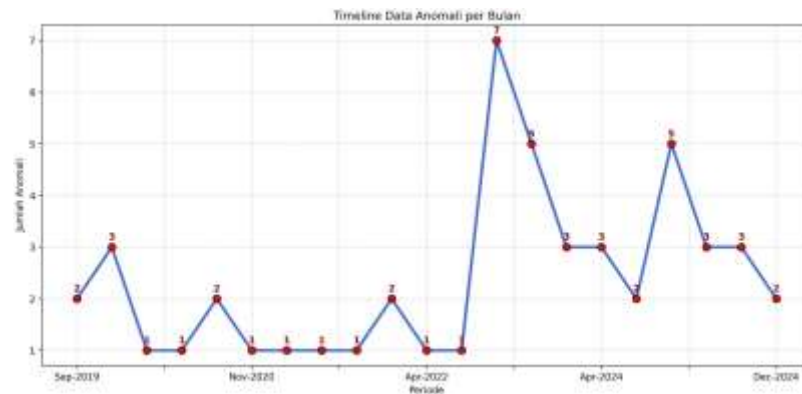
kecamatan dengan jumlah anomali terbanyak memiliki kinerja pelayanan yang lebih buruk.

Berdasarkan KBLI, kode 47722 dan 47721 masing-masing muncul sebanyak sembilan kali pada kelompok anomali. Kode 3226 dan 45407 masing-masing tercatat tiga kali, sementara beberapa kode lainnya muncul dua kali. Temuan ini menunjukkan bahwa observasi anomali terkonsentrasi pada sejumlah kode kegiatan usaha dalam dataset. Akan tetapi, jumlah tersebut belum menunjukkan risiko atau masalah pada sektor usaha tertentu karena distribusi total permohonan per KBLI dan konteks administrasi masing-masing permohonan belum dibandingkan.



Gambar 1. Scatter Plot Hasil Deteksi Anomali Menggunakan Algoritma Isolation Forest

Visualisasi berupa scatter plot, pie chart, boxplot, distribusi anomaly score, distribusi berdasarkan kecamatan dan KBLI, serta timeline digunakan untuk membantu membaca pola hasil. Secara umum, visualisasi menunjukkan bahwa sebagian besar observasi diberi label normal dan sebagian kecil ditandai sebagai anomali.



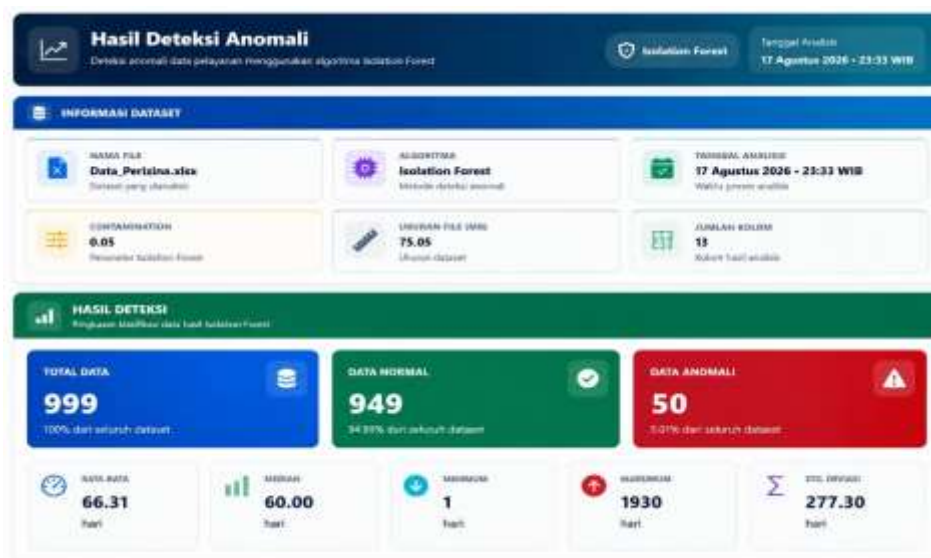
Gambar 2. Timeline Data Anomali

Boxplot juga memperlihatkan adanya nilai durasi yang jauh lebih tinggi dibandingkan sebagian besar observasi. Visualisasi ini bersifat eksploratif: pemisahan visual maupun keberadaan nilai pencilan tidak dengan sendirinya membuktikan terjadinya kesalahan pelayanan.

Halaman Hasil Deteksi

Halaman hasil deteksi digunakan untuk menampilkan hasil penerapan algoritma *Isolation Forest* terhadap dataset yang telah melalui proses *preprocessing*. Pada halaman ini sistem menampilkan informasi mengenai jumlah keseluruhan data, jumlah data normal, jumlah data anomali, serta hasil deteksi masing-masing data.

Berdasarkan hasil penelitian, dari 999 data yang digunakan dalam pemodelan diperoleh 949 data normal atau 95,0% dan 50 data anomali atau 5,0%. Data normal diberikan label 1, sedangkan data yang teridentifikasi sebagai anomali diberikan label -1.



Gambar 3 Halaman Hasil Deteksi Anomali

Hasil tersebut digunakan sebagai informasi awal untuk mengetahui data pelayanan perizinan yang memiliki karakteristik berbeda dari mayoritas data. Data yang teridentifikasi sebagai anomali tidak langsung dinyatakan sebagai

kesalahan pelayanan, tetapi dapat digunakan sebagai bahan untuk pemeriksaan lebih lanjut.

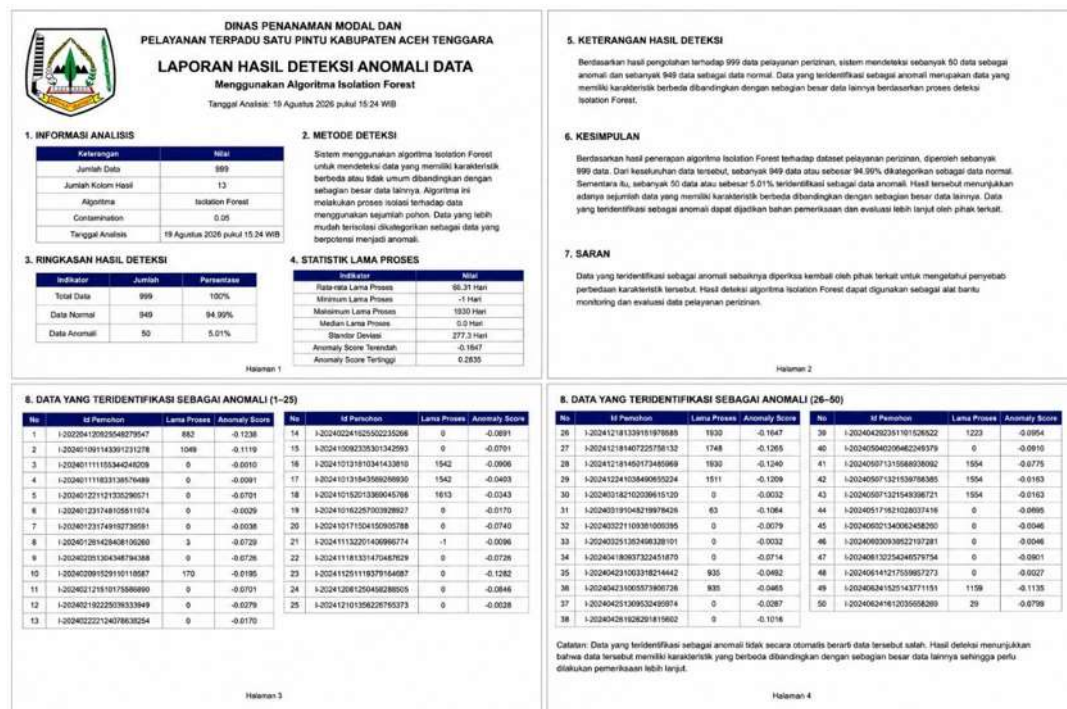
Halaman visualisasi digunakan untuk menyajikan hasil analisis dalam bentuk grafik sehingga informasi hasil deteksi dapat lebih mudah dipahami oleh pengguna. Visualisasi membantu memberikan gambaran mengenai distribusi data normal dan anomali serta karakteristik data yang teridentifikasi sebagai anomali.

Visualisasi yang ditampilkan dalam sistem meliputi grafik distribusi data normal dan anomali, distribusi *anomaly score*, scatter plot, serta visualisasi anomali berdasarkan variabel tertentu seperti Lama Proses, Kecamatan, dan KBLI.



Gambar 4 Halaman Visualisasi Hasil Deteksi

Melalui halaman visualisasi, pengguna dapat melihat pola dan karakteristik data secara lebih informatif dibandingkan hanya menggunakan tabel hasil deteksi. Visualisasi tersebut juga digunakan sebagai pendukung dalam proses analisis dan evaluasi hasil penerapan algoritma *Isolation Forest*.



Gambar 5 Halaman Laporan Hasil Deteksi Anomali

Pada bagian akhir laporan, sistem menyajikan keterangan hasil deteksi, kesimpulan, dan saran. Berdasarkan hasil penelitian, data yang teridentifikasi sebagai anomali dapat dijadikan bahan pemeriksaan dan evaluasi lebih lanjut oleh pihak terkait. Hasil deteksi juga dapat digunakan sebagai alat bantu dalam melakukan monitoring dan evaluasi terhadap data pelayanan perizinan.

Dengan demikian, halaman laporan deteksi menjadi tahap akhir dari implementasi sistem, yang menghubungkan seluruh proses mulai dari unggah dataset, preprocessing, deteksi anomali, hingga penyajian hasil. Halaman ini memastikan hasil penerapan *Isolation Forest* dapat disajikan secara terstruktur, terdokumentasi, dan mudah digunakan sebagai bahan evaluasi data pelayanan perizinan

Pembahasan

Temuan utama penelitian adalah bahwa *Isolation Forest* dapat digunakan sebagai alat penyaringan awal untuk menemukan catatan pelayanan perizinan yang memiliki karakteristik berbeda dari mayoritas data. Dalam konteks administrasi perizinan, pendekatan ini berpotensi membantu petugas memprioritaskan pemeriksaan pada catatan dengan durasi sangat panjang atau kombinasi atribut yang tidak lazim. Hasil deteksi dapat dimanfaatkan untuk menelusuri kembali tanggal penerbitan, status permohonan, jenis izin, klasifikasi usaha, dan informasi pendukung lainnya. Dengan demikian, model lebih tepat ditempatkan sebagai sistem pendukung pemeriksaan daripada sebagai mekanisme otomatis untuk menetapkan kesalahan atau pelanggaran.

Interpretasi hasil perlu mempertimbangkan dua batasan metodologis. Pertama, tidak tersedianya *ground truth* membuat penelitian ini belum dapat menghitung akurasi, precision, recall, atau F1-score, serta belum dapat memastikan bahwa seluruh 50 observasi yang ditandai benar-benar merupakan anomali substantif. Kedua, Label Encoding pada fitur kategorikal nominal dapat memberi kesan urutan numerik yang sebenarnya tidak memiliki makna. Penelitian lanjutan dapat membandingkan strategi representasi fitur yang lebih sesuai, menguji sensitivitas hasil terhadap variasi parameter *contamination* dan *random_state*, serta memvalidasi daftar anomali bersama petugas yang memahami proses bisnis perizinan.

Selain itu, jumlah anomali per kecamatan dan KBLI sebaiknya dilaporkan bersama jumlah keseluruhan permohonan pada masing-masing kelompok agar dapat dihitung proporsi anomali yang lebih bermakna. Rentang tanggal pada catatan yang digunakan juga perlu dijelaskan secara konsisten, sebab dataset disebut mencakup Januari–September 2024, sementara contoh tanggal penerbitan izin dalam hasil pengolahan mencapai Desember 2024 dan tanggal terbit OSS mencakup tahun-tahun sebelumnya. Klarifikasi periode tersebut penting agar cakupan data dan interpretasi durasi pelayanan dapat direplikasi secara tepat.

Secara keseluruhan, model menghasilkan 50 kandidat anomali dari 999 catatan, dengan karakteristik yang antara lain mencakup durasi pelayanan sangat panjang serta sebaran tertentu menurut kecamatan dan KBLI. Hasil ini memberikan dasar untuk pemeriksaan lanjutan terhadap kualitas data dan proses administrasi. Namun, keputusan tindak lanjut tetap perlu didasarkan pada verifikasi data sumber dan konteks pelayanan, bukan semata-mata pada label model.

KESIMPULAN

Berdasarkan hasil penelitian, algoritma *Isolation Forest* berhasil diterapkan untuk mendeteksi anomali pada data pelayanan perizinan di Dinas Penanaman Modal dan Pelayanan Terpadu Satu Pintu (DPMPTSP) Kabupaten Aceh Tenggara melalui sistem berbasis web. Penelitian ini menggunakan 999 data pelayanan perizinan yang telah melalui tahapan *preprocessing*, meliputi pemeriksaan *missing value*, data duplikat, perhitungan atribut Lama Proses, dan transformasi data kategorikal menggunakan *Label Encoding*. Dengan nilai *contamination* sebesar 0,05, algoritma *Isolation Forest* mengidentifikasi 949 data normal (94,99%) dan 50 data anomali (5,01%). Data anomali menunjukkan karakteristik yang berbeda berdasarkan nilai Lama Proses dan *anomaly score*, termasuk durasi pelayanan yang sangat tinggi maupun rendah. Sistem berbasis web yang dikembangkan mampu mengintegrasikan proses pengolahan data, deteksi anomali, visualisasi hasil, dan pembuatan laporan sehingga mendukung pemantauan serta evaluasi data pelayanan perizinan secara lebih sistematis. Meskipun demikian, hasil deteksi anomali merupakan indikasi awal yang memerlukan verifikasi lebih lanjut berdasarkan data sumber dan kondisi pelayanan yang sebenarnya, bukan bukti bahwa data tersebut pasti mengandung kesalahan. Dengan demikian, penerapan algoritma *Isolation Forest* dapat menjadi alat bantu analisis untuk mengidentifikasi

data yang memerlukan pemeriksaan lebih lanjut serta mendukung peningkatan kualitas pengelolaan data dan efektivitas pelayanan perizinan di DPMPTSP Kabupaten Aceh Tenggara.

DAFTAR PUSTAKA

- Aggarwal, C. C. (2024). Unsupervised Learning: A Probabilistic View. In *Probability and Statistics for Machine Learning* (pp. 393–433). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-53282-5_9
- Alang, K. S., & Bindewari, S. (2025). Machine Learning for Anomaly Detection and Prediction in Network Data. *Journal of Quantum Science and Technology*, 2(2). <https://doi.org/10.63345/jqst.v2i2.268>
- Atif, M. (2022). Data mining. *International Journal of Communication and Information Technology*, 3(1), 37–40. <https://doi.org/10.33545/2707661X.2022.v3.i1a.44>
- Buschjäger, S., Honysz, P.-J., & Morik, K. (2022). Randomized outlier detection with trees. *International Journal of Data Science and Analytics*, 13(2), 91–104. <https://doi.org/10.1007/s41060-020-00238-w>
- Calikus, E., Nowaczyk, S., & Dikmen, O. (2025). Context discovery for anomaly detection. *International Journal of Data Science and Analytics*, 19(1), 99–113. <https://doi.org/10.1007/s41060-024-00586-x>
- Cao, Y., Xiang, H., Zhang, H., Zhu, Y., & Ting, K. M. (2025). Anomaly Detection Based on Isolation Mechanisms: A Survey. *Machine Intelligence Research*, 22(5), 849–865. <https://doi.org/10.1007/s11633-025-1554-4>
- Cheng, X., Liu, Z., Tang, M., Du, Y., & Xu, C. (2023a). Anomaly detection based on improved isolated forest. *2023 IEEE 3rd International Conference on Information Technology, Big Data and Artificial Intelligence (ICIBA)*, 971–975. <https://doi.org/10.1109/ICIBA56860.2023.10165081>
- Cheng, X., Liu, Z., Tang, M., Du, Y., & Xu, C. (2023b). Anomaly detection based on improved isolated forest. *2023 IEEE 3rd International Conference on Information Technology, Big Data and Artificial Intelligence (ICIBA)*, 971–975. <https://doi.org/10.1109/ICIBA56860.2023.10165081>
- Choirunnisa, L., Oktaviana, T. H. C., Ridlo, A. A., & Rohmah, E. I. (2023). Peran Sistem Pemerintah Berbasis Elektronik (SPBE) Dalam Meningkatkan Aksesibilitas Pelayanan Publik di Indonesia. *Sosio Yustisia: Jurnal Hukum Dan Perubahan Sosial*, 3(1), 71–95. <https://doi.org/10.15642/sosyus.v3i1.401>
- Dallah, D., Sulieman, H., & Alzaatreh, A. (2024). Outlier Detection Based on Range Statistic. In *Statistical Outliers and Related Topics* (pp. 153–164). CRC Press. <https://doi.org/10.1201/9781003379881-8>
- Foorthuis, R. (2021). On the nature and types of anomalies: a review of deviations in data. *International Journal of Data Science and Analytics*, 12(4), 297–331. <https://doi.org/10.1007/s41060-021-00265-1>
- Hariadi, , Susanti, G., & Rusdi, M. (2025). E-Government Policy Study on Service Quality in Soppeng Regency. *KnE Social Sciences*, 10(18), 1122–1129.

<https://doi.org/10.18502/kss.v10i18.19539>

- He, M., & Chen, X. (2025). Anomaly detection algorithm for big data based on *Isolation Forest* algorithm. *Journal of Computational Methods in Sciences and Engineering*, 25(5), 4143–4154. <https://doi.org/10.1177/14727978251337984>
- Huang, X., Yun, P., & Hu, Z. (2023). A Survey of Collective Anomaly Detection on Sequence Dataset. *International Journal of Data Warehousing and Mining*, 19(1), 1–22. <https://doi.org/10.4018/IJDWM.327363>
- Jain, V., & Tiwari, S. K. (2024). OVERVIEW: MACHINE LEARNING (pp. 130–144). <https://doi.org/10.58532/nbennurch183>
- Jayabharathi, S., & Ilango, V. (2023). *Anomaly Detection Using Machine Learning Techniques: A Systematic Review* (pp. 553–572). https://doi.org/10.1007/978-981-99-3250-4_42
- Kim, K. (2024). Fuzzy anomaly scores for *Isolation Forest*. *Applied Soft Computing*, 166, 112193. <https://doi.org/10.1016/j.asoc.2024.112193>
- Koren, O., Koren, M., & Peretz, O. (2023). A procedure for anomaly detection and analysis. *Engineering Applications of Artificial Intelligence*, 117, 105503. <https://doi.org/10.1016/j.engappai.2022.105503>
- Li, J., Fan, S., & Li, X. (2024). Comparison and Application Research of Unsupervised Anomaly Detection Algorithms. 2024 *International Conference on Artificial Intelligence and Power Systems (AIPS)*, 182–188. <https://doi.org/10.1109/AIPS64124.2024.00047>
- Li, T., Wang, Z., Liu, S., & Lin, W.-Y. (2021). Deep Unsupervised Anomaly Detection. 2021 *IEEE Winter Conference on Applications of Computer Vision (WACV)*, 3635–3644. <https://doi.org/10.1109/WACV48630.2021.00368>
- Liu, T., Zhou, Z., & Yang, L. (2024). Layered *Isolation Forest*: A multi-level subspace algorithm for improving *Isolation Forest*. *Neurocomputing*, 581, 127525. <https://doi.org/10.1016/j.neucom.2024.127525>
- Liu, T., Zhou, Z., & Yang, L. (2024). Layered *Isolation Forest*: A multi-level subspace algorithm for improving *Isolation Forest*. *Neurocomputing*, 581, 127525. <https://doi.org/10.1016/j.neucom.2024.127525>
- Madhuri, G. S., & Usha Rani, M. (2020). *Statistical Approaches to Detect Anomalies* (pp. 499–509). https://doi.org/10.1007/978-981-15-0135-7_46
- Mahalle, P. N., Hujare, P. P., & Shinde, G. R. (2023). *Data Acquisition and Preparation* (pp. 11–38). https://doi.org/10.1007/978-981-99-4850-5_2
- Mahmud, J. S., & Lendak, I. (2024). Enhancing One-Class Anomaly Detection in Unlabeled Datasets Through Unsupervised Data Refinement. 2024 *IEEE 22nd Jubilee International Symposium on Intelligent Systems and Informatics (SISY)*, 000497–000502. <https://doi.org/10.1109/SISY62279.2024.10737577>
- Maia, R., & Antunes, C. (2021). *Density-Based Evaluation Metrics in Unsupervised Anomaly Detection Contexts* (pp. 311–322). https://doi.org/10.1007/978-3-030-85529-1_25
- Masakuna, J. F., Nkashama, D. K., Soltani, A., Frappier, M., Tardif, P. M., Kabanza, F., &

- Kafunda, P. K. (2024). *Impact of Inaccurate Contamination Ratios on Robust Unsupervised Anomaly Detection: Experimental Investigation*. <https://doi.org/10.36227/techrxiv.173396113.31607552/v1>
- Muhammad Ikhsan, Novlianun Dly, Nelly Agustina Siregar, & Maisyarah. (2024). IMPLEMENTASI SISTEM PELAYANAN IZIN USAHA SECARA TERPADU TERHADAP KUALITAS PELAYANAN PUBLIK PADA DPMPSTP KOTA MEDAN. *J-ENSITEC*, 11(01), 10169–10172. <https://doi.org/10.31949/jensitec.v11i01.11926>
- Paluszczek, M., & Thomas, S. (2024). An Overview of Machine Learning. In *MATLAB Machine Learning Recipes* (pp. 1–19). Apress. https://doi.org/10.1007/978-1-4842-9846-6_1
- Qin, Q., Li, H., & Merlo, E. (2026). Unsupervised, robust, and lightweight detection of data pattern anomalies and outliers. *Information and Software Technology*, 192, 108035. <https://doi.org/10.1016/j.infsof.2026.108035>
- Sarifudin, R., & Damanik, F. H. S. (2024). Enhancing transparency and accountability in public administration through information technology utilization. *Gema Wiralodra*, 15(1), 32–40. <https://doi.org/10.31943/gw.v15i1.661>
- Sejr, J. H., & Schneider-Kamp, A. (2021). Explainable outlier detection: What, for Whom and Why? *Machine Learning with Applications*, 6, 100172. <https://doi.org/10.1016/j.mlwa.2021.100172>
- Thakur, A. (2022). A SYSTEMATIC REVIEW ON DATA MINING METHODS AND APPLICATIONS. *International Journal of Advanced Research in Computer Science*, 13(2), 28–31. <https://doi.org/10.26483/ijarcs.v13i2.6807>
- Vishwakarma, P. K., & Jain, N. (2022). A Review of Data Analysis Methodology. 2022 *2nd International Conference on Advance Computing and Innovative Technologies in Engineering* (ICACITE), 356–360. <https://doi.org/10.1109/ICACITE53722.2022.9823843>
- Yepmo, V., Smits, G., Lesot, M.-J., & Pivert, O. (2024). Leveraging an Isolation Forest to Anomaly Detection and Data Clustering. *Data & Knowledge Engineering*, 151, 102302. <https://doi.org/10.1016/j.datak.2024.102302>
- Yusuf, I. M., & Imawan, S. A. (2025). Assessing the performance: e-government in public services in Semarang. *Publisia: Jurnal Ilmu Administrasi Publik*, 10(1), 23–32. <https://doi.org/10.26905/pjiap.v10i1.14024>
- Zadeja, I., & Osis, K. (2025). ANOMALY DETECTION FRAMEWORKS TAXONOMY. SOCIETY. TECHNOLOGY. SOLUTIONS. *Proceedings of the International Scientific Conference*, 3, 69–70. <https://doi.org/10.35363/ViA.sts.2025.142>
- Zhang, L., & Liu, L. (2022). Data Anomaly Detection Based on Isolation Forest Algorithm. 2022 *International Conference on Computation, Big-Data and Engineering* (ICCBE), 87–89. <https://doi.org/10.1109/ICCBE56101.2022.9888169>