

RESEARCH ARTICLE

Deteksi Anomali Data *Time Series* pada Operasional Industri Pipa Gas Alam menggunakan Metode *Bidirectional Long Short-Term Memory* (Bi-LSTM)

M. Taufik Anugrah, Widi Astuti* and Aditya Firman Ihsan

Fakultas Informatika, Universitas Telkom, Bandung, 40257, Jawa Barat, Indonesia

*Corresponding author: widiwdu@telkomuniversity.ac.id

Abstrak

Gas alam merupakan salah satu sumber energi paling penting yang digunakan saat ini, baik oleh rumah tangga, industri, maupun pembangkitan tenaga listrik. Maka menjadi penting untuk menjaga keandalan dari sistem operasional gas alam tersebut. Salah satunya dengan memperhatikan anomali-anomali yang terjadi pada proses distribusi. Penelitian ini bertujuan untuk membuat sebuah sistem deteksi anomali yang terjadi pada operasional pipa gas alam secara *programmatic* dengan memanfaatkan teknologi pembelajaran mesin. Metode yang digunakan adalah *Long Short-Term Memory* (LSTM) dan *Bidirectional Long Short-Term Memory* (Bi-LSTM). Teknik ekstraksi fitur yang digunakan adalah *peak-to-peak anomaly detection*. Penelitian ini mendapatkan nilai akurasi sebesar 99% pada LSTM dan 99% pada Bi-LSTM dengan nilai *loss* sebesar 0,0220.

Key words: Anomali, BiLSTM, Gas Alam, LSTM

Pendahuluan

Gas alam memainkan peran yang sangat penting dan krusial bagi pasokan energi baik di Indonesia maupun di seluruh dunia. Pemanfaatannya tidak terbatas pada sektor industri saja, melainkan juga pada sektor rumah tangga, pertanian hingga pembangkitan tenaga listrik. Kemudahan distribusi, keandalan dan nilai investasi yang dinilai cukup murah dan menguntungkan membuat gas alam lebih dipilih dalam beberapa kasus sebagai sumber energi daripada minyak dan batubara. Menurut data dari Kementerian Energi & Sumber Daya Mineral Republik Indonesia (ESDM), realisasi distribusi gas alam domestik pada tahun 2023 mencapai 3.745 BBTUD. Dimana realisasinya sebanyak 40,5% untuk industri, 18,4% untuk pertanian dan sisanya digunakan untuk kelistrikan serta pasokan LNG dan LPG sektor rumahan. Sisanya dieksport dengan realisasi mencapai angka 1.749 BBTUD. Nilai ini meningkat dari realisasi pemanfaatan gas alam domestik pada tahun 2022 yang hanya bernilai 3.638 BBTUD, namun mengalami penurunan pada sektor ekspor dari angka 1.791 BBTUD [1]. Dari data ini dapat dilihat bahwa gas alam tidak hanya berperan dalam memenuhi pasokan energi dalam negeri, melainkan juga meningkatkan devisa negara melalui skala ekspor.

Latar Belakang

Mengingat pentingnya pasokan gas alam baik bagi industri dalam negeri, pertanian, pembangkitan tenaga listrik dan rumah tangga maka menjadi penting untuk selalu menjaga keandalan dan performa dari proses pendistribusian gas alam tersebut. Hal ini terutama dikarenakan jarak yang cenderung jauh dari pusat produksi ke sub-sub distribusi atau ke pelanggan. Banyak anjungan produksi terletak di lepas pantai yang jauh dari daratan, dengan jaringan pipa berkilo-kilometer yang menghubungkannya ke fasilitas penerima terpusat di darat. Hal ini menjadi tantangan tersendiri dalam proses distribusi dan transmisi lepas pantai. Stabilitas gas alam yang diterima pun harus dijaga sedemikian rupa, sehingga setiap platform sumber dapat mengkoordinasikan keseimbangan tekanan gas yang ditransmisikan dengan tetap menjaga kualitas dan pasokan energi terpenuhi [2]. Salah satu teknik yang dapat dilakukan adalah dengan menganalisis anomali-anomali yang terjadi dan membuat sebuah model *preventive maintenance* pada sistem distribusi gas alam. Hal ini bertujuan untuk melakukan deteksi dini dan meminimalisir dampak dari kemungkinan-kemungkinan anomali yang terjadi [3].

Anomali adalah sebuah titik data atau kelompok titik data yang menunjukkan penyimpangan signifikan dari titik atau kelompok tetangganya dan tidak sesuai dengan tren keseluruhan data [4]. Deteksi anomali mengacu pada proses mengidentifikasi kejadian, objek, atau pengamatan yang tidak biasa yang menyimpang secara substansial

dari perilaku atau pola khas [5]. Secara umum, anomali dapat didekripsi secara visual dari pengamatan-pengamatan dan pemahaman yang mendalam dari data-data instrumen pengukuran. Namun teknik ini sarat akan human error, dan saat anomali terjadi, maka sudah terlambat untuk melakukan mitigasi. Pendekatan lain yang lebih mutakhir dapat dilakukan dengan membuat sebuah model deteksi anomali dengan memanfaatkan teknologi pembelajaran mesin [6]. Salah satu metode yang paling populer adalah *Long Short-Term Memory* (LSTM) dan *Bidirectional Long Short-Term Memory* (Bi-LSTM). Pendekatan berbasis BiLSTM menggunakan dua jaringan LSTM searah untuk memproses data temporal. Satu jaringan LSTM memproses data dari masa lalu ke masa depan, sementara yang lain memproses data dari masa depan ke masa lalu. Pendekatan ini meningkatkan efisiensi proses ekstraksi fitur dari data temporal, sehingga menghasilkan performa deteksi anomali yang lebih baik [7]. Penelitian ini mencoba membuat sebuah sistem deteksi anomali pada data operasional industri pipa gas alam dengan metode LSTM & Bi-LSTM. Penelitian ini diharapkan dapat memberikan kontribusi untuk meningkatkan efisiensi dan keandalan operasional industri pipa gas alam serta meminimalisir dampak akibat adanya anomali.

Perumusan Masalah

Berdasarkan pemaparan latar belakang masalah, maka rumusan masalah yang ingin diangkat pada penelitian ini adalah sebagai berikut:

1. Bagaimana mengimplementasikan metode LSTM & BiLSTM untuk membuat sebuah sistem deteksi anomali pada data operasional industri pipa gas alam dengan metode ekstraksi fitur *peak-to-peak anomaly detection*?
2. Bagaimana performa metode LSTM & Bi-LSTM dalam mendekteksi anomali yang terjadi pada data sistem operasional pipa gas alam dengan metode ekstraksi fitur *peak-to-peak anomaly detection*?

Tujuan

Berdasarkan pemaparan pada bagian latar belakang dan rumusan masalah, maka tujuan dari penelitian ini dapat dirumuskan sebagai berikut:

1. Membuat sebuah model sistem deteksi anomali pada data operasional industri pipa gas alam menggunakan metode LSTM & Bi-LSTM dengan metode ekstraksi fitur *peak-to-peak anomaly detection*.
2. Mengukur performansi model sistem deteksi anomali pada data operasional industri pipa gas alam menggunakan metode LSTM & Bi-LSTM dengan metode ekstraksi fitur *peak-to-peak anomaly detection*.

Organisasi Tulisan

Tulisan ini secara garis besar mendeskripsikan temuan dan hasil analisis terhadap temuan pada penelitian model sistem deteksi anomali pada data operasional industri pipa gas alam menggunakan metode LSTM & Bi-LSTM dengan metode ekstraksi fitur *peak-to-peak anomaly detection*. Adapun bagian-bagian yang akan dipaparkan pada penulisan Tugas Akhir ini adalah sebagai berikut:

1. Pendahuluan

Bagian ini berisi latar belakang dan urgensi penelitian, perumusan masalah yang diangkat pada penelitian ini serta tujuan atau luaran yang hendak dicapai.

2. Studi Terkait

Bagian berisi pemaparan studi literatur dan landasan teori yang relevan dengan penelitian.

3. Sistem yang Dibangun

Bagian ini akan berisi tentang penjelasan desain penelitian, metodologi yang digunakan, model yang dibuat, serta skenario pengujian.

4. Evaluasi

Bagian ini akan berisi pemaparan hasil dan analisis dari pemaparan hasil yang didapat selama proses penelitian berlangsung.

5. Kesimpulan

Bagian ini akan berisi hasil dan kesimpulan secara garis besar dari penelitian yang sudah dilakukan.

Tinjauan Pustaka

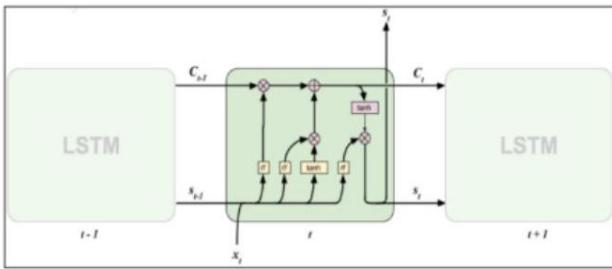
Penelitian Terkait

Salah satu penelitian terkait dilakukan oleh A.S Raihan & I. Ahmed pada Tahun 2023. Penelitian ini diterapkan pada kumpulan data time series multivariant yang dikumpulkan dari Pembangkit Listrik Tenaga Angin (PLTA). Model Bi-LSTM Autoencoder mencapai akurasi klasifikasi sebesar 96,79% [8]. Pada tahun 2023 Lee dkk. menggunakan metode *peak-to-peak anomaly detection* pada penelitiannya. Metode ini menggunakan pendekatan *local maximum* pada beberapa basis data yang terbagi dalam rentang waktu tertentu. Penilaian sebuah anomali dilakukan berdasarkan nilai ambang batas yang didasarkan pada sebuah data terhadap kelompok tetangganya [9]. Pada tahun 2022 Aljameel dkk [6]. melakukan penelitian deteksi anomali pada dataset pengukuran pipa migas yang memiliki 8 fitur. Pada penelitian ini digunakan 5 jenis metode pembelajaran mesin yang berbeda, yaitu *random forest*, *Support Vector Machine* (SVM), KNN, *gradient boosting*, dan *decision tree*. Hasil dari pengujian 5 metode tersebut adalah didapat bahwa metode SVM dapat memperoleh akurasi terbesar dibandingkan 4 metode yang lain, yaitu sebesar 97,43% [6].

Lalu ada penelitian pada tahun 2021 S. Aljbali & K. Roy juga melakukan perbandingan performansi model LSTM dan Bi-LSTM terhadap dataset UNSW-NB15 yaitu berisi paket jaringan yang melibatkan serangkaian fitur dari *payload* dan *header* paket dan mendapatkan hasil model prediksi dengan menggunakan Bi-LSTM dengan tingkat akurasi mencapai 99,70% [10]. Pada tahun 2021 Pooja & Purohit melakukan penelitian dengan hasil analisis akurasi pada dataset KDDCUP-99 menggunakan Bi-LSTM yang paling tinggi yaitu 0,997 [11]. Pada tahun 2020 D. Pan dkk. melakukan penelitian yang mendapatkan hasil percobaan membuktikan bahwa model Bi-LSTM yang diusulkan dapat meningkatkan efisiensi pemantauan. Akurasi dari algoritma RNN, LSTM, dan BiLSTM menunjukkan Bi-LSTM paling tinggi yaitu 0,969 pada dataset1, 0,976 pada dataset2, dan 0,957 pada dataset [12].

Deteksi Anomali

Anomali adalah sebuah fenomena dalam sebuah data yang menunjukkan indikasi ketidaknormalan jika dibandingkan dengan trend keseluruhan data atau pada kelompok data tetangganya [4]. Deteksi anomali mengacu pada proses mengidentifikasi kejadian, objek, atau pengamatan yang tidak biasa yang secara substansial menyimpang dari perilaku atau pola tipikal. Hal ini terutama karena peristiwa yang tidak biasa dapat memberikan wawasan berharga terhadap fenomena yang sedang dipelajari [13]. Dalam beberapa kasus, anomali adalah sebuah data rusak yang bisa berupa *spark* atau *outlier* yang terlalu tinggi [8]. Namun tidak semua *spark* atau *outlier* yang terlalu tinggi dapat dikategorikan sebagai sebuah anomali. Contohnya adalah *spark* yang muncul pada gelombang listrik pada saat sebuah alat baru dinyalakan. *Spark* ini hanya muncul sesaat saja dan dapat dijelaskan menggunakan fenomena laplace. Terminologi anomali lebih merujuk kepada *spark* yang tidak biasa muncul pada keadaan tertentu [14]. Dalam membuat sebuah model *machine learning*, data yang bersifat anomali adalah data kebetulan yang tidak masuk kedalam area normal. Keberadaan data anomali bukan hanya merusak trend dan meningkatkan *loss factor* secara keseluruhan, namun juga membuat model menjadi bias. Sehingga dikhawatirkan memberikan keputusan yang salah. Maka dalam



Gambar 1. Memory Cell LSTM [8]

beberapa kasus, anomali dihapus atau tidak dimasukan kedalam data yang melibatkan pembelajaran mesin. Dalam hal ini, kemampuan pendekripsi anomali menjadi penting. Terutama jika menyangkut situasi *real-time* [15].

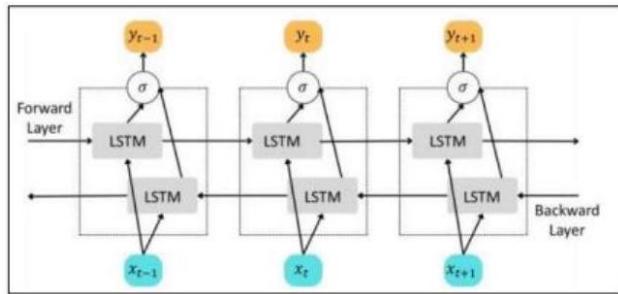
Teknik yang umum dalam melakukan deteksi anomali adalah dengan melakukan deteksi *outlier* pada data. Namun teknik ini harus berhati-hati. Dikarenakan tidak semua *outlier* adalah anomali [16]. Teknik lain yang biasa digunakan adalah *peak analysis* dan *peak-to-peak analysis*. *Peak analysis* mencoba menganalisis data secara global dengan menentukan global maksimum, sementara *peak-to-peak* dilakukan pada kelompok data tertentu dengan menentukan lokal maksimum [9]. Kedua teknik ini mencoba menjadi nilai yang wajar dari kelompok data tertentu hingga pada batas tertentu. Biasanya nilainya adalah 0.9999 dari puncak yang wajar. Penentuan puncak yang wajar ini bisa dilakukan melalui beberapa teknik, misal dengan menggunakan pendekatan deret, atau menentukan secara langsung. Teknik *peak-to-peak* lebih baik jika digunakan pada data yang bersifat *time-series*. Hal ini dikarenakan teknik ini membagi data menjadi beberapa segmen berdasarkan rentang waktu, sehingga dapat lebih mudah dilakukan pendekripsi. Mengingat bisa jadi sebuah anomali yang terjadi pada rentang waktu tertentu dianggap sebagai data biasa secara keseluruhan [17]. Pendekatan lain yang mirip adalah *multi-sector mapping* yang membagi data menjadi beberapa bagian untuk menganalisis anomali dan menyatukannya kembali dalam suatu basis data kolektif [18].

Cross Validation

Metode *Cross Validation* digunakan untuk mempercepat waktu komputasi dalam membangun model, sambil tetap memastikan hasil estimasi model tetap akurat. *Cross validation* atau rotasi estimasi adalah suatu teknik validasi digunakan untuk mengevaluasi hasil analisis. Teknik *cross validation* umumnya digunakan untuk memprediksi tingkat keakuratan suatu model prediktif dan salah satu teknik *cross validation* dimana data dibagi menjadi K bagian set data dengan jumlah data yang sama. Metode *K-Fold Cross Validation* digunakan untuk memperkirakan kesalahan prediksi dan mengevaluasi model dengan melakukan beberapa pengujian pada data yang diminta oleh mesin. Dari pembagian *fold* ini bisa dibagikan menjadi 5 sampai dengan 10 *fold* tergantung kebutuhan untuk dijadikan sebagai *data testing* dan *data training* sebagai model evaluasi dari setiap *fold* yang dihasilkan oleh metode *K-fold cross validation* [19].

Long Short-Term Memory (LSTM)

Long short-term memory (LSTM) merupakan sebuah evolusi dari arsitektur RNN. Algoritma LSTM termasuk ke dalam *deep neural network architecture* dimana dalam satu langkah hasil dari output akan dimasukan kembali kedalam sistem sehingga sistem membentuk perulangan. Algoritma LSTM memiliki *memory cell* dan *gate units* pada setiap neurons nya yang berfungsi untuk mengatur memori dalam setiap neuron [8]. Alur dari *memory cell* LSTM diilustrasikan pada Gambar 1. Pada



Gambar 2. Arsitektur Bi-LSTM [8]

Gambar 1 terlihat bagaimana alur kerja dari cell algoritma LSTM. Terdapat empat tahapan yang menggunakan *activation function* dalam setiap input yaitu *forget gates*, *input gates*, *cell gates* dan *output gates*. Terlihat bagaimana beberapa cell LSTM yang saling berhubungan. Alur pemrosesan data dimulai dengan mengambil nilai c_t dan s_t dari cell sebelumnya, jika sebelumnya tidak ada cell atau cell pertama maka nilai c_t dan s_t akan dibuat menjadi 0. Setelah itu nilai s_t akan diproses bersama dengan input yang masuk kedalam cell LSTM melalui *forget gate*, *input gate* dan *output gate* hingga menghasilkan s_t . Dan nilai c_t akan diproses di dalam *forget gate* dan digabungkan dengan hasil perhitungan *input gate* dan *output gate* hingga menghasilkan c_t [20].

Bidirectional Long Short-Term Memory (Bi-LSTM)

Bi-LSTM terdiri dari LSTM ke depan dan LSTM ke belakang. Karena *track outlier* tidak terjadi secara terpisah, struktur maju dan mundur model Bi-LSTM lebih cocok untuk mendekripsi *track outlier*. Unit jaringan saraf LSTM adalah unit dasar yang membentuk jaringan saraf Bi-LSTM [21]. Lapisan LSTM terdiri dari rangkaian sel LSTM, dan data masukan hanya diproses dalam arah maju. Di sisi lain, Bi-LSTM menyertakan lapisan LSTM tambahan yang memproses data dalam arah mundur, seperti yang ditunjukkan pada Gambar 2. Melalih jaringan Bi-LSTM setara dengan melalih dua jaringan LSTM searah yang terpisah. Salah satu jaringan ini dilatih berdasarkan urutan masukan asli, sementara jaringan lainnya dilatih berdasarkan salinan urutan masukan yang dibalik. Pendekatan ini memberikan jaringan informasi yang lebih kontekstual, yang mengarah pada pembelajaran masalah yang lebih cepat dan komprehensif [8].

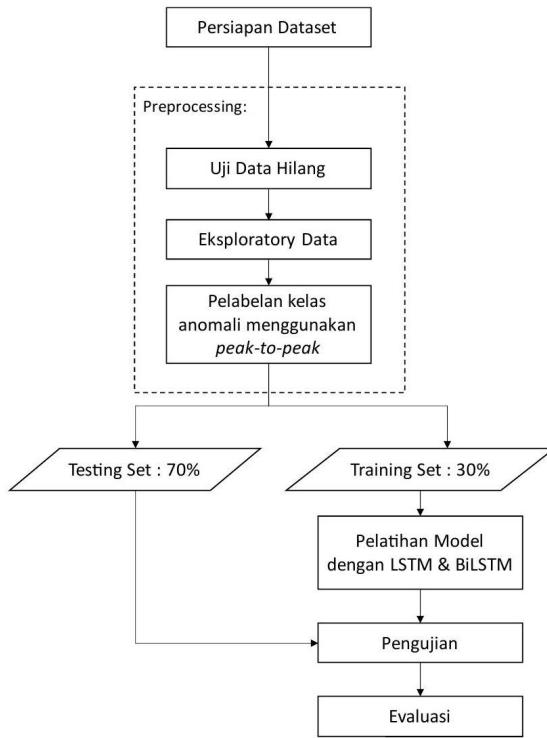
Metodologi Penelitian

Alur Penelitian

Langkah-langkah yang dilakukan pada penelitian ini dapat dilihat pada diagram alir pada Gambar 3. Penelitian dimulai dengan mempersiapkan dataset. Dataset yang sudah disiapkan kemudian masuk kedalam proses *pre-processing* dimana data akan dibersihkan. Kemudian data akan dianalisis untuk mengetahui *trend data* dan kemudian dilakukan proses pendekripsi kelas anomali menggunakan *peak-to-peak anomaly detection*. Setelah itu dilakukan pelabelan kelas anomali dari hasil yang didapatkan dari *peak-to-peak*. Kemudian data dibagi menjadi *data training* dan *data testing* dengan perbandingan 70:30. Lalu dilanjutkan dengan pembuatan model LSTM dan Bi-LSTM dan dilakukan pengujian serta evaluasi akhir.

Dataset

Dataset yang digunakan dalam penelitian ini sebagai masukan model Bi-LSTM yang diterapkan adalah hasil *data record* dari data operasional transmisi pipa gas perusahaan minyak dan gas yang berlokasi



Gambar 3. Alur Penelitian

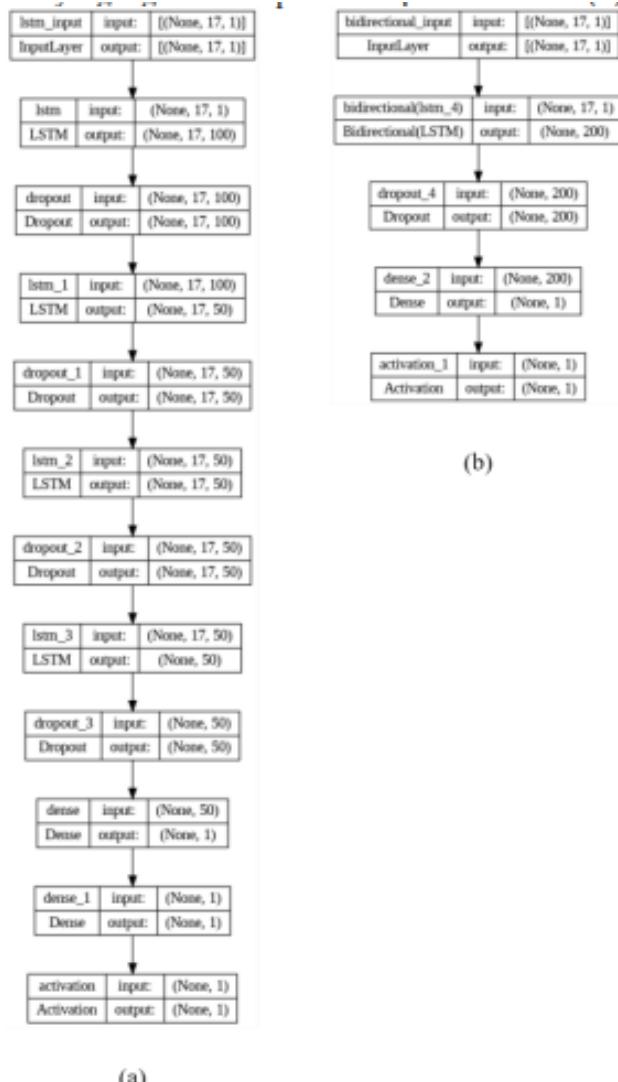
di laut Natuna. Dataset ini berisi *record* data dari tahun 2013 sampai tahun 2021 yang masing-masing berisi sekitar 29.000 *data record* operasi dengan 20 buah kolom.

Arsitektur Model

Model yang digunakan pada penelitian ini adalah *Long Short-Term Memory* (LSTM) dan *Bidirectional Long Short-Term Memory* (Bi-LSTM). Arsitektur LSTM yang digunakan memiliki empat buah layer LSTM dengan jumlah unit masing-masing 100, 50, 50 dan 50. Setelah masing-masing layer LSTM kemudian diisi dengan *layer drop out* sebesar 50%. *Layer* terakhir adalah *dense layer* dengan jumlah 1 unit sebanyak dua *layer*. Dan kemudian ditutup dengan *activation function* ReLu. Sementara arsitektur Bi-LSTM dibuat lebih sederhana dengan menggunakan satu buah *layer* Bidirectional LSTM dengan jumlah 100 unit. Kemudian ditambahkan *layer dropout* setelahnya sebesar 50% dan terakhir adalah *dense function* sebanyak 1 unit. Kemudian arsitektur ditutup dengan *activation function* ReLu. Arsitektur LSTM yang digunakan dapat dilihat pada Gambar 4 (a), sementara arsitektur Bi-LSTM yang digunakan dapat dilihat pada Gambar 4 (b). Proses pelatihan model dilakukan dengan menggunakan 5 kali iterasi pada arsitektur Gambar 4. *Optimizer* yang digunakan adalah Adam. Metric yang digunakan selama proses *training* adalah *loss* dan akurasi.

Metric Pengujian

Metric pengujian paling utama yang digunakan adalah *confusion matrix*. *Confusion matrix* adalah sebuah matrix yang menunjukkan perbandingan antara keputusan yang diprediksi oleh model dengan data yang sebenarnya. *Confusion matrix* memberikan kita 4 buah data, yaitu *True Positive* / TP (data positif yang berhasil diprediksi), *True Negative* / TN (data negatif yang berhasil diprediksi), *False Negative* / FN (data positif yang gagal diprediksi), dan *False Positive* / FP (data negatif



Gambar 4. Arsitektur LSTM (a), Arsitektur BiLSTM (b)

		Classifier Prediction	
		Positive	Negative
Actual Value	Positive	True Positive	False Negative
	Negative	False Positive	True Negative

Gambar 5. Confusion Matrix

yang gagal diprediksi). Dari *confusion matrix* inilah nanti akan didapatkan beberapa metric pengujian lain seperti: akurasi, *precision*, *recall* dan *f1-score*. Contoh dari *confusion matrix* dapat dilihat pada Gambar 5. Metrik pengujian pertama yang digunakan dalam penelitian ini adalah akurasi. Akurasi menyatakan seberapa akurat model memprediksi data apakah positif, negatif, atau netral. Untuk menghitung nilai akurasi

matriks konfusi dapat digunakan Rumus (1).

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

Metrik selanjutnya yang digunakan dalam penelitian ini adalah *precision*. *Precision* membandingkan jumlah data positif yang berhasil diprediksi dengan jumlah data yang tergolong positif. Untuk menghitung presisi matriks konfusi dapat digunakan Rumus (2).

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

Metrik selanjutnya yang digunakan dalam penelitian ini adalah *recall*. *Recall* merupakan perbandingan antara jumlah data positif yang berhasil diprediksi dibagi dengan jumlah data yang benar-benar positif. Penarikan kembali menyatakan seberapa baik model memprediksi data positif. Untuk menghitung *recall* dari matriks konfusi dapat digunakan Rumus (3).

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

Metric terakhir yang digunakan adalah *F1-Score*. *F1-Score* adalah sebuah metrik yang membandingkan hasil kali antara *recall* dan *precision* dengan hasil penjumlahan antara *recall* dan *precision*. Rumus yang digunakan dapat dilihat pada Rumus (4).

$$F1 \text{ Score} = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}} \quad (4)$$

Hasil dan Pembahasan

Bagian ini akan menjelaskan tentang deskripsi hasil dari eksperimen yang dilakukan selama proses penelitian berlangsung. Analisis akan dimulai dengan memaparkan data hasil penelitian, pembersihan *outlier*, pendekstian data-data anomali menggunakan *peak-to-peak anomaly detection* dan terakhir akan dipaparkan perbandingan performansi antara *Long Short-Term Memory* (LSTM) dan *Bidirectional Long Short Term Memory* (Bi-LSTM) dalam pendekstian anomali.

Uji Data Hilang

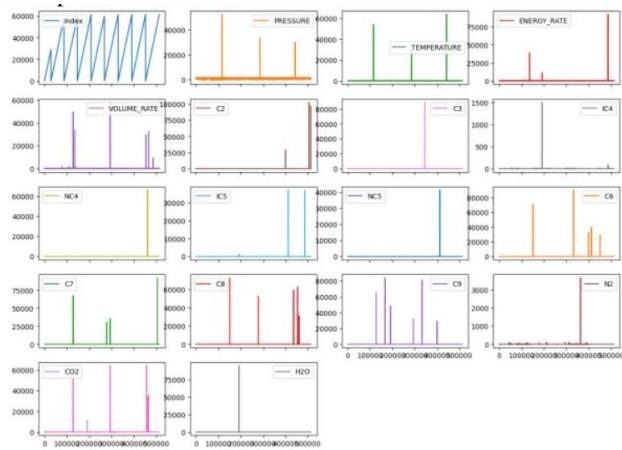
Analisis yang pertama dilakukan adalah analisis uji data hilang. Analisis ini bertujuan untuk menghapus data-data yang hilang serta mengeliminasi data yang mayoritas hilang. Dari total 21 data yang dimiliki oleh dataset, 2 data dieliminasi sebagai data hilang dan 2 data lainnya dieleminasi sebagai data yang tidak penting. Sehingga data yang berhasil diseleksi dan digunakan dalam pemrosesan berikutnya berjumlah 17 data, antara lain: *pressure*, *temperature*, *energy rate*, *volume rate*, C1, C2, C3, IC4, IC5, NC5, C6, C7, C8, C9, N2, CO2 dan H2O.

Eksploratory Data Analysis

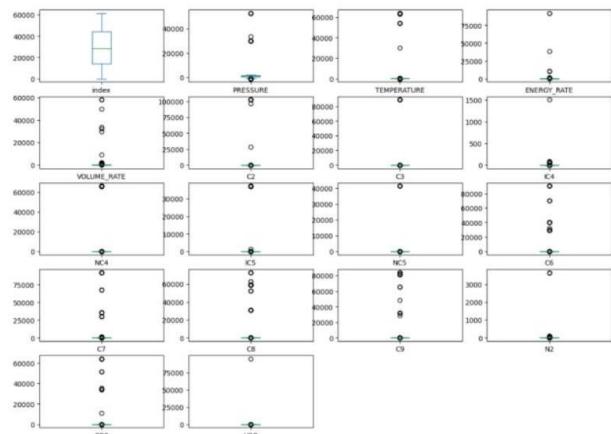
Analisis EDA dimulai dengan melakukan visualisasi data dari data-data yang berhasil diseleksi oleh uji data hilang. Adapun trend data yang berhasil diseleksi dapat dilihat pada Gambar 6. Dari hasil visualisasi data pada Gambar 6, ditemukan bahwa pada setiap data memiliki *spark* yang sangat-sangat tinggi dan sangat tidak beraturan. Hal ini hampir tidak mungkin dihasilkan kecuali karena faktor kesalahan (error). Data-data ini kemudian diklasifikasikan sebagai *outlier* (pencikan) yang akan dieliminasi dan tidak disertakan pada pemrosesan berikutnya.

Outlier Removal

Outlier removal adalah sebuah teknik pembersihan *outlier* yang digunakan sebagai akibat atau implikasi langsung dari temuan data-data ekstrim pada proses EDA. Proses ini merupakan lanjutan dari proses



Gambar 6. Visualisasi Trend Dataset



Gambar 7. Box Plot Outlier Detection

pre-processing dimana data-data yang bersifat *outlier* atau didefinisikan sebagai *outlier* secara statistik akan dieliminasi. Metode yang dilakukan untuk mendefinisikan *outlier* adalah box plot. Hasil dari visualisasi box plot dapat dilihat pada Gambar 7. Dari hasil visualisasi box plot pada Gambar 7 dapat dilihat bahwa hampir keseluruhan data memiliki *outlier* yang nilainya sangat-sangat jauh dari *baseline* median. *Outlier* ini kemudian dibersihkan hingga pada batasan tertentu sampai garis *baseline* yang menyatakan median secara visual terlihat pada *box plot*. Artinya data tidak melenceng jauh dari median.

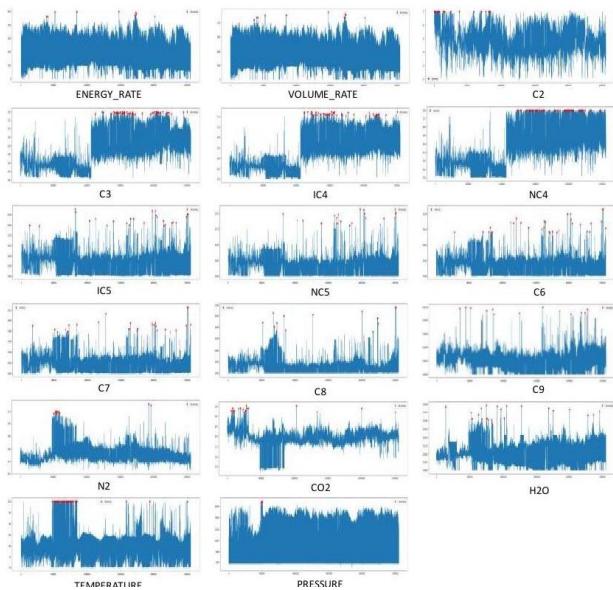
Proses ini memerlukan eksperimen *trial and error* karena sebuah data terkait dengan data yang lain, sehingga pembersihan *outlier* yang terlalu rendah akan menghasilkan *outlier* baru pada data lainnya. Oleh karena itu nilai yang dijadikan batasan data untuk membersihkan *outlier* ini tidak didefinisikan secara umum, melainkan spesifik pada setiap data. Pembersihan dilakukan hingga garis median berada ditengah atau setidaknya terlihat pada diagram box plot. Setelah melalui proses *outlier removal*, kemudian terseleksi sebanyak 254846 data yang sudah bersih dan dapat digunakan pada pemrosesan berikutnya. Sampel dataset bersih dapat dilihat pada Tabel 1.

Anomaly Detection

Proses deteksi anomali bertujuan untuk mendekstikan anomali yang ada pada data sebagai data latih untuk pembelajaran *machine learning*. Pendekstian anomali dilakukan pada semua data untuk membentuk

Table 1. Sampel Data Bersih

Index	Pressure	Temperature	Energy Rate	Volume Rate	...	H ₂ O
0	1001, 58	83,4	107, 92	97,8	...	0,00 88
1	590,1 9	76,65	73,2	67,6	...	0,00 69
2	579,7 2	76.03	325, 6	300,5	...	0,00 88
...
2548	589,7	82,71	63,7	59,4	...	0,02 08

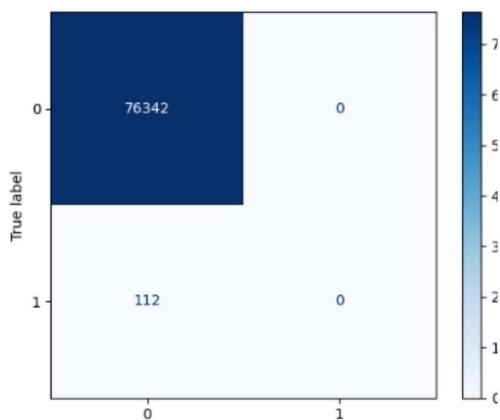
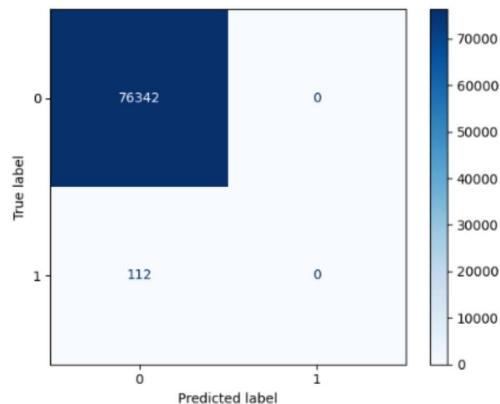
**Gambar 8.** Hasil Deteksi Anomali

kelas-kelas anomali. Metode yang digunakan adalah *peak-to-peak anomaly detection*. Cara kerjanya yaitu dengan menentukan jarak minimum antara dua puncak yang terdeteksi. Dalam hal ini, puncak harus berjarak setidaknya 500 sampel. Lalu memastikan bahwa input merupakan array 1D dan menggunakan *local maxima* serta menggunakan kuantil 0.9995 untuk menetapkan batas puncak. Hanya puncak yang melebihi batas yang dianggap *anomaly*. Hasil dari deteksi anomali pada keseluruhan data dapat dilihat pada Gambar 8. Data yang didalamnya terdapat indikasi anomali, kemudian secara kolektif diberi label '1'. Meskipun hanya terdapat satu anomali didalamnya. Sementara data yang didalamnya tidak terdapat indikasi anomali, secara kolektif diberi label '0'. Dari hasil pengujian, ditemukan 365 data yang memiliki indikasi *anomaly* dari total 254846 data. Hasil dari deteksi anomali pada keseluruhan data dapat dilihat pada Tabel 2.

Hasil LSTM

Proses *training* pada model LSTM dilakukan menggunakan *optimizer* Adam pada 5 buah iterasi (*epoch*). Hasil yang didapatkan pada setiap iterasi dapat dilihat pada Tabel 3, sementara hasil pengujian LSTM dapat dilihat pada Tabel 4. *Confusion Matrix* hasil pengujian LSTM dapat dilihat pada Gambar 9.

Untuk hasil *cross validation* dari LSTM sebanyak 5 *fold* dapat dilihat pada Tabel 5.

**Gambar 9.** *Confusion Matrix* LSTM**Gambar 10.** *Confusion Matrix* Bi-LSTM

Hasil Bi-LSTM

Proses *training* pada model Bi-LSTM dilakukan menggunakan *optimizer* Adam pada 5 buah iterasi (*epoch*). Hasil yang didapatkan pada setiap iterasi dapat dilihat pada Tabel 6, sementara hasil pengujian Bi-LSTM dapat dilihat pada Tabel 7. *Confusion Matrix* hasil pengujian Bi-LSTM dapat dilihat pada Gambar 10.

Untuk hasil *cross validation* dari Bi-LSTM sebanyak 5 *fold* dapat dilihat pada Tabel 8.

Analisis

Dari Tabel 3 dapat dilihat bahwa nilai *loss*, *validation loss*, akurasi, dan akurasi data validasi pada model LSTM cenderung tidak memiliki pergerakan. Dari iterasi pertama hingga iterasi akhir didapatkan nilai *loss* sebesar 0.0220, *loss* validasi sebesar 0.0224, akurasi sebesar 0.9986 dan nilai akurasi pada data validasi sebesar 0.9985. Hal ini bukan berarti model tidak mengalami kemajuan dalam proses pelatihan, melainkan hasil pelatihan pada iterasi pertama sudah memiliki hasil yang terbaik dengan akurasi mencapai 99%. Analisis lain adalah adanya jumlah data yang tidak imbang antara data yang memiliki anomali dan data yang tidak memiliki anomali. Hal ini mengakibatkan kecondongan ekstrim yang membuat model cenderung pintar pada salah satu kelas saja dan ketidakmampuan mendeteksi kelas yang lain. Hal ini dibuktikan dari hasil *confusion matrix* yang menghasilkan nilai hasil prediksi cenderung pada label '0' atau tidak memiliki anomali. Hal ini membuat nilai *precision*, *recall*, dan *F1-Score* pada label '0' mencapai 100% sementara nilai *precision*, *recall* dan *F1-Score* pada label '1'

Table 2. Deteksi Anomali

Index	Pressure	Temperature	Energy Rate	Volume Rate	...	H ₂ O	Anomaly
0	1001,58	83,4	107,92	97,8	...	0,0 088	0
1	590,19	76,65	73,2	67,6	...	0,0 069	0
2	579,72	76,03	325,6	300,5	...	0,0 088	0
3	580,00	76,23	324,36	299,76	...	0,0 101	0
254842	1259,63	73,26	127,16	116,75	...	0,0 049	0
254843	988,22	85,07	216,49	201,98	...	0,0 201	0
254844	579,26	81,78	216,49	201,98	...	0,0 201	0
254845	589,78	82,71	63,7	59,4	...	0,0 208	0

Table 3. Hasil Iterasi LSTM

Iterasi	Loss	Val Loss	Akurasi	Val Akurasi
1	0,0220	0,0224	0.9986	0.9985
2	0,0220	0,0224	0.9986	0.9985
3	0,0220	0,0224	0.9986	0.9985
4	0,0220	0,0224	0.9986	0.9985
5	0,0220	0,0224	0.9986	0.9985

Table 4. Hasil Pengujian LSTM

Label	Precision	Recall	F1-Score
0	100%	100%	100%
1	0%	0%	0%

Table 5. Cross Validation LSTM

Fold	Label	Precision	Recall	F1Score
1	0	100%	100%	100%
	1	0%	0%	0%
2	0	100%	100%	100%
	1	0%	0%	0%
3	0	100%	100%	100%
	1	0%	0%	0%
4	0	100%	100%	100%
	1	0%	0%	0%
5	0	100%	100%	100%
	1	0%	0%	0%

Table 6. Hasil Iterasi Bi-LSTM

Iterasi	Loss	Val Loss	Akurasi	Val Akurasi
1	0,0241	0,0224	0.9982	0.9985
2	0,0220	0,0224	0.9986	0.9985
3	0,0220	0,0224	0.9986	0.9985
4	0,0220	0,0224	0.9986	0.9985
5	0,0220	0,0224	0.9986	0.9985

hanya sebesar 0%. Pada Tabel 6 hasil yang sedikit berbeda ditunjukkan dimana pembelajaran cenderung lebih terlihat terjadi pada model

Table 7. Hasil Pengujian Bi-LSTM

Label	Precision	Recall	F1-Score
0	100%	100%	100%
1	0%	0%	0%

Table 8. Cross Validation Bi-LSTM

Fold	Label	Precision	Recall	F1-Score
1	0	100%	100%	100%
	1	0%	0%	0%
2	0	100%	100%	100%
	1	0%	0%	0%
3	0	100%	100%	100%
	1	0%	0%	0%
4	0	100%	100%	100%
	1	0%	0%	0%
5	0	100%	100%	100%
	1	0%	0%	0%

Bi-LSTM. Pada model ini, didapatkan nilai *loss* yang menurun pada iterasi kedua dan akurasi yang cenderung meningkat pada iterasi kedua. Namun setelahnya, metric pengujian yang dihasilkan tidak berubah. Pada iterasi pertama didapatkan nilai *loss* sebesar 0.0241, *validation loss* sebesar 0.0224, akurasi 0.9982 dan akurasi pada data validasi sebesar 0.9985. Pada iterasi kedua dan seterusnya didapatkan nilai *loss* sebesar 0.0220 (menurun 0.0021 dari iterasi pertama), *validation loss* sebesar 0.0224, akurasi sebesar 0.9986 (naik 0.0004 dari iterasi pertama) dan akurasi pada data validasi sebesar 0.9985.

Nilai metrik performansi yang cenderung tetap pada iterasi kedua dan seterusnya bukan berarti model tidak mengalami pembelajaran, melainkan model sudah mencapai performanya yang paling baik. Hal ini ditunjukkan dengan nilai akurasi yang mencapai 99%. Analisis lainnya sama dengan model LSTM dimana adanya jumlah data yang tidak imbang antara data yang memiliki anomali dan data yang tidak memiliki anomali. Ketidakseimbangan ini mengakibatkan kecondongan ekstrim yang membuat model cenderung pintar pada salah satu kelas saja dan ketidakmampuan mendeteksi kelas yang lain. *Confusion matrix* yang menunjukkan nilai hasil prediksi cenderung pada label '0' atau tidak memiliki anomali. Akibatnya nilai *precision*, *recall*, dan *F1-Score* pada label '0' mencapai 100% sementara nilai *precision*, *recall* dan *F1-Score* pada label '1' hanya sebesar 0%. Meskipun sudah dilakukan *cross validation* sebanyak 5 *fold* menunjukkan nilai cenderung pada label '0' atau tidak memiliki anomali dikarenakan jumlah data yang masih tidak

Table 9. Perbandingan LSTM dan Bi-LSTM

Iterasi	LSTM		Bi-LSTM	
	Loss	Akurasi	Loss	Akurasi
1	0,0220	0.9986	0,0241	0.9982
2	0,0220	0.9986	0,0220	0.9986
3	0,0220	0.9986	0,0220	0.9986
4	0,0220	0.9986	0,0220	0.9986
5	0,0220	0.9986	0,0220	0.9986

imbang atau terlalu jauh perbedaannya antara data anomali dan data tidak anomali. Pada *fold 1* menggunakan *train set* sebanyak 203876 dan *test set* sebanyak 50970 masih mendapatkan hasil 100% pada label '0' untuk *F1-Score*. Sedangkan pada *fold 2*, *fold 3*, *fold 4* dan *fold 5* menggunakan *train set* sebanyak 203877 dan *test set* sebanyak 50969 mendapatkan hasil 100% pada label '0'.

Dari analisis diatas, dapat ditemukan perbedaan antara model LSTM dan model Bi-LSTM. Data dengan model LSTM langsung menemui performa paling maksimal, sedangkan model Bi-LSTM harus melalui satu iterasi pembelajaran untuk menemukan hasil yang maksimal seperti yang ditampilkan pada Tabel 9. Perbedaan ini dapat dianalisis dari faktor kerumitan arsitektur model. Jika dibandingkan antara arsitektur model LSTM dan arsitektur model Bi-LSTM dapat dilihat bahwa arsitektur LSTM memiliki tingkat kerumitan dan kompleksitas yang lebih tinggi dibandingkan dengan arsitektur Bi-LSTM. Dimana model LSTM yang dibuat menggunakan 4 buah layer LSTM, sementara Bi-LSTM dibuat hanya menggunakan satu buah layer. Hal ini juga yang dapat menjadi pertimbangan mengapa model LSTM cenderung lebih cepat menemukan hasil yang terbaik ketimbang model Bi-LSTM.

Kesimpulan

Kesimpulan yang didapatkan dari hasil penelitian Deteksi Anomali Data *Time Series* pada Operasional Industri Pipa Gas Alam Menggunakan Metode LSTM dan Bi-LSTM adalah sebagai berikut:

1. *Anomaly detection* secara statistik dapat dideteksi dengan menggunakan metode *peak-to-peak anomaly detection*. Namun membutuhkan waktu dan pemrosesan yang relatif panjang. Hal ini dapat dipersingkat dengan pendekatan pembelajaran mesin dengan menggunakan algoritma LSTM dan Bi-LSTM.
2. Model LSTM memiliki kinerja yang lebih baik daripada Bi-LSTM karena mendapatkan performa terbaik dalam waktu yang relatif lebih singkat, yaitu pada iterasi pertama.
3. Kompleksitas arsitektur yang tepat dapat menghasilkan proses pembelajaran yang lebih cepat untuk menemui performa yang paling baik.
4. Pada iterasi terakhir, model LSTM dan Bi-LSTM cenderung menghasilkan nilai performa yang sama dimana akurasi yang dihasilkan mencapai 99% dan loss yang dihasilkan sebesar 0.0220.

Daftar Pustaka

1. Kementerian ESDM RI. Capaian Kinerja Sektor Esdm Tahun 2023 Dan Target 2024; 2024.
2. Ihsam A, Darmadi, Uttungadewa S, Rahmawati S, Giovanni I, Himawan S. MultiLayer LSTM Implementation in Operational Condition Forecasting of a Natural Gas Transmission Pipeline Network. In: Proceedings of the 2022 International Conference on Information and Communications Technology (ICOIACT); 2022. p. 244-9.
3. Mohamed M, Lazim H, Osman W. Preventive Maintenance: Some Operations and Technology Related Practices at a Malaysian Oil and Gas Firm. Journal of Advanced Research in Applied Sciences and Engineering Technology. 2023;31:1-18.
4. Ahmed I, Dagnino A, Ding Y. Unsupervised Anomaly Detection Based on Minimum Spanning Tree Approximated Distance Measures and its Application to Hydropower Turbines. IEEE Transactions on Automation Science and Engineering. 2019;16(2):654-67.
5. Elsayed M, Le-Khac NA, Dev S, Jurcut A. Network Anomaly Detection Using LSTM Based Autoencoder; 2020.
6. Aljameel S, et al. An Anomaly Detection Model for Oil and Gas Pipelines Using Machine Learning. Computation. 2022;10(8):138.
7. Lee S, et al. Smart Metering System Capable of Anomaly Detection by Bi-directional LSTM Autoencoder. In: Proceedings of the IEEE International Conference on Consumer Electronics (ICCE); 2022. .
8. Raihan A, Ahmed I. A Bi-LSTM Autoencoder Framework for Anomaly Detection – A Case Study of a Wind Power Dataset; 2023.
9. Lee BS, Kaufmann JC, Rizzo DM, Haq IU. Peak Anomaly Detection from Environmental Sensor-Generated Watershed Time Series Data. In: Information Management and Big Data; 2023. p. 142-57.
10. Aljballi S, Roy K. Anomaly Detection Using Bidirectional LSTM. In: Advances in Computational Intelligence Systems; 2021. p. 612-9.
11. TS P, Shrinivasacharya P. Evaluating neural networks using Bi-Directional LSTM for network IDS (intrusion detection systems) in cyber security. Global Transitions Proceedings. 2021;2(2):448-54.
12. Pan D, Song Z, Nie L, Wang B. Satellite Telemetry Data Anomaly Detection Using BiLSTM Prediction Based Model. In: Proceedings of the 2020 IEEE International Instrumentation and Measurement Technology Conference (I2MTC); 2020. p. 1-6.
13. Duan X, Fu Y, Wang K. Data Preprocessing Technology in Network Traffic Anomaly Detection. In: Proceedings of the International Conference on Artificial Intelligence and Cloud Computing; 2023. p. 227-35.
14. Feng L, et al. Anomaly detection for electricity consumption in cloud computing: framework, methods, applications, and challenges. EURASIP Journal on Wireless Communications and Networking. 2020;2020.
15. Barszcz T, Zabarylo M. Fault detection method based on an automated operating envelope during transient states for the large turbomachinery. Journal of Vibroengineering. 2022;24.
16. Divya D, Marath B, Kumar MB. Review of fault detection techniques for predictive maintenance. Journal of Quality in Maintenance Engineering. 2022;29.
17. Lis A, Dworakowski Z, Czubak P. An anomaly detection method for rotating machinery monitoring based on the most representative data. Journal of Vibroengineering. 2021;23.
18. Mallioris P, Aivazidou E, Bechtis D. Predictive maintenance in Industry 4.0: A systematic multi-sector mapping. CIRP Journal of Manufacturing Science and Technology. 2024;50:80-103.
19. Firmansyach WA, Hayati U, Wijaya YA. Analisa Terjadinya Overfitting dan Underfitting pada Algoritma Naive Bayes dan Decision Tree dengan Teknik Cross Validation. JATI (Jurnal Mahasiswa Teknik Informatika). 2023;7(1):262-9.
20. Choir TB. Multivariate Time Series Anomaly Detection Dengan Metode Long Short Term Memory Untuk Validasi Data Kelembaban Tanah [Undergraduate Thesis]; 2022.
21. Cao Y, Cao J, Zhou Z, Liu Z. Aircraft Track Anomaly Detection Based on MOD-Bi-LSTM. Electronics. 2021;10(9).