Deteksi Anomali Lalu Lintas Jaringan Internal Inbound Dan Outbound Menggunakan Algoritma Long Short-Term Memory

1st Khairunnisa Salsabila Riswanti
Fakultas Rekayasa Industri
Universitas Telkom
Bandung, Indonesia
khairunnisasalsabila@student.telkomuniversity.ac.id

2nd Faqih Hamami
Fakultas Rekayasa Industri
Universitas Telkom
Bandung, Indonesia
faqihhamami@telkomuniversity.ac.id

3rd Tien Fabrianti Kusumasari
Fakultas Rekayasa Industri
Universitas Telkom
Bandung, Indonesia
tienkusumasari@telkomuniversity.ac.id

***Abstrak*— Saat ini penggunaan internet sudah menjadi kebutuhan dalam kegiatan sehari-hari. Berdasarkan laporan DataReportal, pengguna internet di Indonesia pada Januari 2022 ada sebanyak 73,7%. Data tersebut menunjukan bahwa seiring berkembangnya era digital, pengguna internet juga akan terus bertambah. Setiap aktivitas penggunaan internet akan terekam dalam suatu lalu lintas jaringan inbound dan outbound. Pada lalu lintas jaringan inbound dan outbound, akan menampilkan tren data normal. Namun dapat juga muncul data yang diluar tren yang disebut sebagai data anomali. Lalu lintas jaringan anomali tersebut dapat terjadi karena adanya peningkatan yang signifikan dalam volume data lalu lintas jaringan. Anomali pada data lalu lintas jaringan inbound dan outbound juga terjadi pada data lalu lintas jaringan PT XYZ yang merupakan perusahaan yang berfokus pada bidang jasa layanan TIK dan jaringan telekomunikasi di Indonesia. Untuk mencegah terjadinya data anomali, dapat menggunakan IDS melalui deteksi anomali dengan algoritma yang dapat memproses data sekuen dan data skala besar. Algoritma yang digunakan dalam penelitian ini adalah LSTM. Penelitian ini menggunakan metodologi CRISP-DM sebagai sistematika penyelesaian masalah. Terdapat beberapa tahapan yang diterapkan yaitu business understanding, data understanding, data preparation, modelling, dan evaluasi. Pengujian model dan evaluasi model dilakukan berdasarkan parameter yang ditentukan menghasilkan model yang dapat mendeteksi anomali.**

***Kata Kunci* — deteksi anomali, deep learning, lstm**

I. PENDAHULUAN

Saat ini penggunaan internet sudah menjadi kebutuhan dalam kegiatan sehari-hari. Berdasarkan laporan DataReportal, pengguna internet di Indonesia pada Januari 2022 ada sebanyak 73,7% dari total populasi atau sekitar 204,7 juta pengguna. Data tersebut menunjukan bahwa seiring berkembangnya era digital, pengguna internet juga akan terus bertambah. Setiap aktivitas penggunaan internet akan terekam dalam suatu lalu lintas jaringan inbound dan outbound. Lalu lintas jaringan inbound merupakan data jaringan masuk atau unduh yang dilakukan oleh pengguna internet. Sedangkan lalu lintas jaringan outbound merupakan data jaringan keluar atau unggah yang dilakukan oleh pengguna internet. Pada lalu lintas jaringan inbound dan outbound, akan menampilkan tren data normal. Namun dapat juga muncul data yang diluar tren tersebut yaitu data yang berada diatas tren normal atau dapat disebut sebagai data anomali. Lalu lintas jaringan anomali tersebut dapat terjadi karena adanya peningkatan yang signifikan dalam volume data lalu lintas jaringan. Salah satunya dapat diakibatkan masalah jaringan.

Lalu lintas jaringan yang anomali ditunjukan pada perubahan secara tiba-tiba pada grafik multi router traffic grapher (MRTG) yang awalnya konsisten normal pada pergerakannya, tiba-tiba menjadi naik volumenya yang berarti terjadi sesuatu yang tidak biasa sehingga menimbulkan kecurigaan adanya gangguan yang terjadi. Untuk menghindari hal tersebut maka perlu mendeteksi anomali pada lalu lintas jaringan inbound dan outbound supaya dapat mengetahui penyebab adanya data anomali tersebut sehingga dapat melakukan perbaikan layanan untuk memperkecil kemungkinan terjadinya anomali.

Salah satu organisasi yang melakukan monitoring lalu lintas jaringan di Indonesia adalah PT XYZ. PT XYZ merupakan perusahaan yang berfokus pada bidang jasa layanan teknologi informasi dan komunikasi (TIK) dan jaringan telekomunikasi di Indonesia. Salah satu misi PT XYZ adalah mengorkestrasi ekosistem digital untuk memberikan pengalaman digital pelanggan terbaik. Berkaitan dengan misi tersebut, PT XYZ terus melakukan memperbaiki kinerja menjadi lebih baik. Salah satu hal yang harus dilakukan adalah mencegah terjadinya serangan kejahatan cyber dengan menggunakan intrusion detection system (IDS) melalui deteksi anomali. Hal tersebut dilakukan untuk menjaga aktivitas jaringan tetap normal dan aman dengan meminimalisir adanya anomali. Sehingga perusahaan dapat memberikan pengalaman digital pelanggan terbaik tanpa ada gangguan. IDS merupakan sistem deteksi gangguan lalu lintas jaringan yang dapat mencegah terjadinya serangan kejahatan cyber [1]

Penelitian ini ditujuan untuk menganalisis rancangan deteksi anomali lalu lintas jaringan internal inbound dan outbound. Dalam sistematika penyelesaian masalah ini menggunakan algoritma LSTM untuk mendapatkan model yang bisa digunakan untuk mendeteksi anomali lalu lintas jaringan data internal inbound dan outbound. Pada penelitian ini menggunakan sistematika penyelesaian masalah dalam bentuk metodologi CRISP-DM.

Metodologi CRISP-DM atau Cross Industry Standard Process-Data Mining merupakan metodologi yang disusun untuk membuat katalog serta pemandu langkah pengerjaan paling umum dalam proyek penambangan data. CRISP-DM merupakan standar de facto dalam pengembangan proyek penambangan data dan penemuan pengetahuan [2]

II. KAJIAN TEORI

A. Data Anomali Jaringan

Data anomali merupakan data yang tidak normal atau berbeda dari yang lain pada umumnya. Menurut KBBI anomali merupakan ketidaknormalan, penyimpangan dari normal, dan kelainan. Dapat diartikan bahwa data anomali adalah kelainan pada data jaringan yang berada diluar aktivitas normal sehingga menimbulkan kecurigaan atau keanehan yang terjadi berdasarkan anomali yang terjadi.

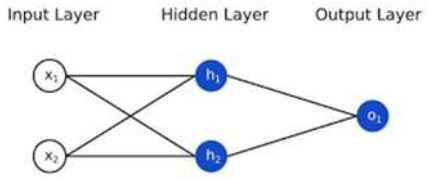
Untuk mengetahui data anomali tersebut maka perlu adanya monitoring atau deteksi pada anomali yang terjadi. Dalam mendeteksi anomali tersebut dapat dilakukan dengan mempelajari bagaimana tren data yang normal dan dicari apa ada data yang menyimpang dari tren data yang normal berdasarkan tren data yang telah dipelajari sebelumnya. Sehingga dapat diketahui mana saja data anomali. Deteksi anomali merupakan suatu kegiatan untuk menganalisis data normal yang digunakan untuk mendeteksi data yang tidak normal [3]

Pada lalu lintas jaringan *inbound* dan *outbound* dapat muncul aktivitas yang tidak biasa atau anomali. Lalu lintas jaringan yang anomali tersebut muncul dikarenakan naiknya frekuensi aktivitas secara drastis. Aktivitas yang menyebabkan adanya anomali pada lalu lintas jaringan bukan hanya karena naiknya frekuensi aktivitas secara drastis saja, dapat juga dikarenakan oleh adanya kerusakan sistem pada komputer pengguna internet ataupun adanya gangguan internet. Hal tersebut dapat menyebabkan performa jaringan dan performa layanan menjadi turun. Anomali pada jaringan dapat terjadi dalam kurun waktu singkat hingga satu jam. Data anomali dapat ditentukan dengan membandingkan tren data satu jam sebelumnya dan tren data satu jam setelahnya. Deteksi anomali pada jaringan dilakukan untuk menghindari performa jaringan dan performa layanan menjadi turun serta mencegah terjadinya data anomali yang lain. Faktor terjadinya data anomali yaitu dikarenakan terdapat perubahan data dari tren data normal [4]

B. Neural Network

Neural Network (NN) merupakan sebuah model komputasi yang terinspirasi dari cara kerja jaringan saraf yang ada pada otak manusia yang dimana informasi akan dialirkan menuju satu titik yaitu neuron dan akan menghasilkan sebuah *output* atau keluaran. Tiap neuron memiliki nilai bobot (*weight*) yang dapat diambil dari hasil perhitungan antara nilai *input* dan nilai *scalar* yang telah diberikan fungsi aktivasi. Kumpulan neuron yang ada pada *neural network* disebut dengan *hidden layer*. Setiap *hidden layer* saling terintegrasi dengan *hidden layer* yang lainnya dan *output layer*. Pada Gambar 1.1 terdapat contoh skema *neural network* yang terdapat nilai x_1 dan x_2 sebagai layer *input* yang kemudian akan diproses dalam *hidden layer* menjadi h_1 dan h_2 yang akan diproses dengan jumlah kemungkinan proses terbanyak yang kemudian akan menghasilkan *output layer* atau hasil sejumlah 1 yaitu o_1 sebagai hasil terbaik dari proses yang sudah dilakukan.

Deep learning merupakan salah satu teknik neural network yang mengelola neuron dalam banyak layer. Sesuai dengan namanya yaitu "*deep*" yang dimana akan melakukan proses secara mendalam. *Deep learning* melakukan pembelajaran lebih dalam menggunakan *neural network* pada layer secara berturut-turut untuk melakukan pembelajaran dari data secara *iterative*



GAMBAR II.1
Skema Neural Network

Untuk menggunakan *deep learning* diperlukan fungsi aktivasi dalam pemodelan. Fungsi aktivasi merupakan fungsi yang digunakan didalam neural network yang berfungsi untuk mengubah sinyal input menjadi sinyal *output*. Cara kerja dari fungsi aktivasi adalah dengan mencari nilai dari hasil perhitungan antara nilai *scalar* dan nilai *input* yang di beri fungsi identitas setelah itu baru menentukan apakah nilai tersebut perlu aktif atau tidak pada neuron. Nilai aktif pada neuron setelah proses perhitungan dilakukan dan sudah diberikan sebuah fungsi aktivasi disebut sebagai bobot.

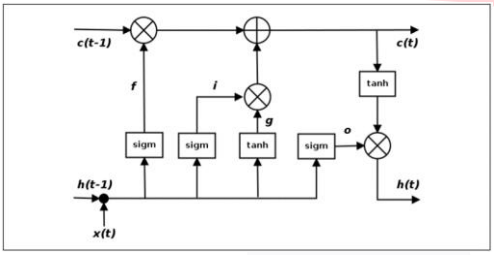
Terdapat banyak jenis fungsi aktivasi, seperti fungsi aktivasi sigmoid, fungsi aktivasi tan hiperbolik (Tanh), dan fungsi aktivasi rectified linear unit RelU. Peneliti menggunakan fungsi aktivasi RelU dikarenakan fungsi aktivasi tersebut merupakan fungsi aktivasi yang banyak digunakan pada neural network dibandingkan dengan fungsi aktivasi sigmoid dan tanh serta mampu melakukan proses komputasi dengan kompleksitas training yang multilayer (Boob dkk., 2022). Fungsi aktivasi RelU yang digunakan untuk mengambil nilai maksimum antara dua nilai yaitu nilai 0 dan x (input) dan juga jenis aktivasi ini umum digunakan karena proses komputasinya yang irit. RelU membantu proses neural network ke GPU lebih awal.

C. Recurrent Neural Network (RNN)

Salah satu algoritma dari *deep learning* yaitu *recurrent neural network* (RNN). RNN merupakan algoritma yang bekerja lebih luas dalam melakukan pembelajaran sesuai dengan namanya yaitu *neural network* yang berarti seperti jaringan saraf yang memiliki banyak cabang. RNN banyak digunakan dalam mesin penerjemah, pengenalan suara, dan bidang lain yang berkaitan dengan *natural language processing* (NLP). RNN dapat memproses data sekuensial seperti kata dalam kalimat, catatan not musik, dan sejenisnya yang menampilkan ketergantungan pada setiap elemen sebelumnya. RNN sel dapat menggabungkan ketergantungan tersebut dengan menyembunyikan keadaan tersebut atau memorinya yang menahan esensi sebenarnya yang terlihat sebelumnya.

D. Long Short-Term Memory (LSTM)

Long short-term memory (LSTM) merupakan salah satu algoritma varian RNN yang dapat *mempelajari long-term dependencies*. LSTM dapat melakukan proses komputasi dengan baik pada berbagai macam permasalahan yang ada. Selain itu LSTM merupakan varian RNN yang paling banyak digunakan. Pada RNN dapat menggabungkan status tersembunyi dari langkah waktu sebelumnya dan *input* pada saat ini menggunakan layer tanh untuk dapat mengimplementasikan perulangan. LSTM juga dapat menggunakan perulangan dengan cara yang sama, dengan perbedaan jumlah layer yaitu bukan satu layer tanh seperti pada RNN, tetapi terdapat empat layer yang dapat berinteraksi dengan menggunakan cara yang sangat spesifik. Pada Gambar II.2 menggambarkan arsitektur umum dari proses komputasi LSTM. Arsitektur tersebut menampilkan bagaimana transformasi yang diterapkan pada *hidden layer* dengan *time step* t. Garis pada bagian atas merupakan *cell* c yang merepresentasikan memori unit internal. Garis pada bagian bawah merupakan *hidden layer* h. Kemudian *gate* i, f, o, dan g merupakan bagian mekanisme yang digunakan LSTM untuk mengatasi permasalahan *vanishing gradient*.



GAMBAR II.2
Arsitektur LSTM

Dalam LSTM terdapat salah satu teknik yaitu *Autoencoder*. *Autoencoder* merupakan salah satu *artificial neural network* yang berfokus untuk mempelajari bagaimana skema *encoding-decoding* data yang paling baik dan efisien. *Autoencoder* terdiri dari *input layer, output layer, encoder, decoder,* dan *space latent*. Data akan dimasukan pada layer neuron pada bagian *encoding* yang dimana data akan dikompres menuju *space latent*. Kemudian pada bagian *decoder* akan menguraikan data yang dikompres pada bagian *encoding* lalu akan digunakan pada bagian *output layer*. *Output* yang sudah melalui proses *encoding-decoding* akan dibandingkan dengan data awal kemudian nilai *error* akan digunakan untuk memperbarui nilai bobot. Pada *autoencoder* data dilatih supaya dapat meminimalisir *nilai reconstruction error* (Nguyen dkk., 2021). Hal tersebut dilakukan supaya dapat membuat *output vector* memiliki nilai yang mirip dengan nilai aslinya. Pada bagian *decoder*, data awal yang digunakan adalah data berdasarkan hasil *output* dari bagian *encoding*.

E. Data Time Series

Data time-series atau data deret waktu merupakan suatu kumpulan data yang memiliki keterikatan dengan data waktu. Hal tersebut memiliki definisi data tersebut tetapi selain itu memiliki riwayat yang berkaitan dengan masa lalu. Data time-series merupakan suatu hal yang cukup penting yang dimana data hasil dari penelitian terhadap data internal inbound dan outbound yang digunakan dari masa lalu yang kemudian akan di kumpulkan kembali dan akan melakukan analisis supaya dapat untuk melihat suatu keterkaitan atau hubungan ataupun korelasi antar data tersebut.

Data *time series* merupakan sesuatu yang cukup penting untuk diteliti dan analisis, kerena pada data time series data yang merupakan hasil dari observasi masa lalu (*historical*) akan dikumpulkan lalu dianalisis. Kemudian setelah dianalisis data tersebut dapat digunakan untuk memprediksi evolusi dinamika sistem dan teknik yang memiliki kompleksitas tinggi. Tujuannya yaitu supaya dapat melakukan peramalan atau forecasting tren masa depan secara akurat berdasarkan data time series dari masa lalu dan saat ini.

F. Mean Absolute Error (MAE)

Mean absolute error (MAE) merupakan salah satu metrik evaluasi yang dapat digunakan. Metrik evaluasi sendiri merupakan metrik yang digunakan untuk mengukur kesesuaian data eksisting dengan data prediksi (Azmi dkk., 2020). MAE merupakan salah satu metrik evaluasi model yang digunakan dengan model regresi. MAE model berhubungan dengan set test data merupakan nilai rata-rata absolut dari kesalahan prediksi individu terhadap semua *instance* dalam set tes data. Setiap kesalahan prediksi merupakan selisih antara nilai sebenarnya dan nilai prediksi untuk *instance*.

MAE merupakan salah satu indikator yang dapat digunakan untuk mengevaluasi kinerja model, yang menentukan rata-rata jarak antara hasil sebenarnya dan prediksi. MAE memiliki kelebihan dibandingkan dengan metrik evaluasi lain seperti MSE karena lebih mudah untuk dipahami dan tidak memberikan dampak signifikan dari kesalahan tertentu. MAE juga digunakan untuk memantau kinerja dari model regresi dan untuk mengevaluasi seberapa akurat model memprediksi target. Berikut merupakan rumus dari MAE.

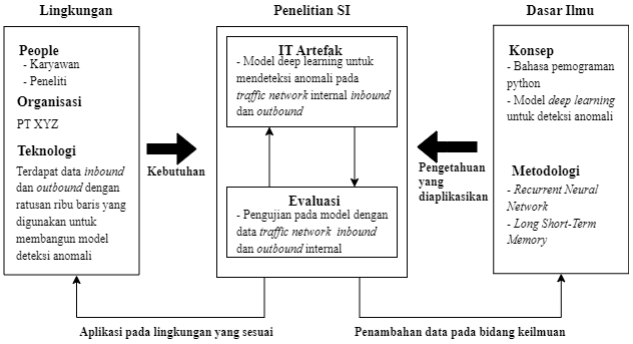
$$mae = \frac{\sum_{i=1}^{n} abs\,(y_i - \lambda(x_i))}{n}$$

Keterangan :
y_i = nilai target sebenarnya untuk nilai x_i
$\lambda(x_i)$ = nilai target yang diprediksi untuk instance uji x_i
n = jumlah instance uji.

III. METODE

A. Kerangka Berpikir

Dalam penelitian ini akan meneliti mengenai pendeteksian anomali lalu lintas jaringan internal *inbound* dan *outbound* pada MRTG di PT XYZ menggunakan algoritma *Recurrent Neural Network* (RNN). Penelitian ini menggunakan model konseptual yang dapat dilihat pada Gambar III.1. Model konseptual merupakan model yang dirancang secara terstruktur untuk mendapatkan dasar pemikiran dari pemecahan masalah.



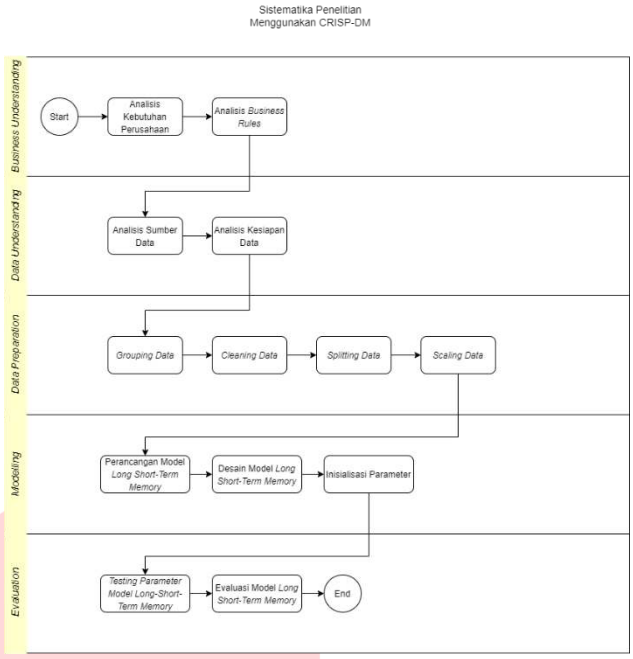
GAMBAR II.1
Kerangka Berpikir

Berdasarkan Gambar III.1 model konseptual yang digambarkan menjelaskan aspek-aspek terkait dalam melakukan penelitian. Terdapat 3 elemen yang berkaitan yaitu lingkungan, penelitian sistem informasi, dan dasar ilmu. Dalam setiap elemen tersebut memiliki komponen masing-masing. Penelitian ini akan melakukan pemodelan menggunakan algoritma LSTM pada data yang telah didapatkan dari perusahaan. Penelitian dapat dikatakan berhasil jika menghasilkan model yang dapat mendeteksi anomali lalu lintas jaringan internal *inbound* dan *outbound*.

B. Sistematika Penyelesaian Masalah

Sistematika penyelesaian masalah merupakan langkah-langkah yang akan dilakukan oleh peneliti dalam mendapatkan solusi untuk mengatasi suatu permasalahan yang mungkin terjadi. Dalam sistematika penyelesaian masalah ini menggunakan algoritma LSTM untuk mendapatkan model yang bisa digunakan untuk mendeteksi anomali lalu lintas jaringan data internal *inbound* dan *outbound*. Pada penelitian ini menggunakan sistematika penyelesaian masalah dalam bentuk metodologi CRISP-DM yang dapat dilihat pada Gambar III.2.

Pada Gambar III.2 menjelaskan mengenai bagaimana sistematika penyelesaian masalah yang akan dilakukan oleh peneliti. Fase pertama yang dilakukan adalah *business understanding* yaitu membahas mengenai penjelasan permasalahan dan kebutuhan perusahaan yang akan diteliti dalam penelitian ini. Kemudian membahas mengenai business rules yang perlu diterapkan dalam pelaksanaan penelitian. Fase selanjutnya yaitu fase *data understanding* yang dimana melakukan analisis sumber data yang diberikan oleh perusahaan melalui wawancara dan kesepakatan. Selain itu pada fase ini juga melakukan analisis kesiapan dari data yang telah diberikan oleh perusahaan. Selanjutnya untuk fase ketiga yaitu fase *data preparation*. Pada fase *data preparation* ini akan dilakukan *grouping* data untuk melakukan analisis berdasarkan id. Selain itu pada fase ini juga akan dilakukan *cleaning data* untuk melakukan pembersihan data agar pada fase *modelling* dapat menghasilkan model yang baik.



GAMBAR II.2
Sistematika Penelitian Menggunakan CRISP-DM

C. Fase Business Understanding

Pada fase ini peneliti melakukan dua proses atau langkah yang dilakukan, yaitu analisis kebutuhan perusahaan dan analisis *business rules* yang dibutuhkan oleh perusahaan. Analisis kebutuhan perusahaan dilakukan untuk memahami kebutuhan dari perusahaan PT XYZ serta melakukan pemahaman permasalahan yang harus diselesaikan dalam penelitian ini. Sedangkan analisis *business rules* dilakukan untuk memahami aturan-aturan yang ditetapkan oleh perusahaan sebagai acuan dalam pelaksanaan penelitian ini.

Proses pertama yang dilakukan yaitu analisis kebutuhan perusahaan. PT XYZ mengelola dan memantau semua lalu lintas jaringan berdasarkan layanan yang sudah disediakan. Perusahaan menggunakan *network interface* sebagai komponen yang memuat lalu lintas jaringan. *Network interface* merupakan komponen yang terdapat dalam server supaya dapat terhubung dengan jaringan internet. Perusahaan menggunakan aplikasi MRTG untuk memantau beban dan lalu lintas jaringan internal. Aplikasi tersebut menampilkan grafik jumlah muatan data *inbound* dan *outbound* secara time-series setiap lima menit pada setiap hari berdasarkan *network interface* yang dijadikan id. Selain itu, aplikasi ini juga memuat kapasitas masing-masing *network interface* untuk *inbound* dan *outbound*.

Setiap id memiliki data *inbound* dan *outbound* yang digunakan setiap lima menit. Pada aktivitas *inbound* dan *outbound* yang bernilai diluar tren normal data atau data anomali akan masuk dalam data tanpa diketahui jika data tersebut merupakan data anomali. Data *inbound* dan *outbound* yang masuk setiap lima menit tersebut akan menjadi 288 data per hari untuk setiap id. Maka dapat disimpulkan data anomali akan menjadi tersembunyi karena total jumlah data inbound dan outbound dari semua id setiap harinya cukup banyak. Sedangkan proses *quality control* dilakukan per satu minggu sekali. Menyikapi hal ini, perusahaan perlu melakukan deteksi anomali data *inbound* dan *outbound* agar dapat mengetahui pada id mana dan kapan adanya data yang anomali. Sehingga dapat

menganalisis penyebab anomali dan perbaikan performa supaya dapat memperkecil kemungkinan terjadi lagi data anomali.

Dengan adanya permasalahan tersebut, maka peneliti melakukan penelitian ini dengan tujuan untuk membantu menyelesaikan permasalahan yang ada. Pada penelitian ini, akan dilakukan deteksi data anomali terhadap *data inbound* dan *outbound* internal pada rentang waktu Januari hingga November 2022. Berdasarkan hasil wawancara mengenai data dengan pihak perusahaan, terdapat parameter data *inbound* dan *outbound* akan dianggap sebagai data anomali yaitu jika terdapat data yang berada diluar tren data normal.

Untuk membuat model deteksi anomali pada data *inbound* dan *outbound* internal perusahaan PT XYZ, terdapat beberapa *business rules* yang harus diikuti berdasarkan wawancara yang telah dilakukan dengan pihak perusahaan. Aturan-aturan tersebut ditujukan untuk membuat batasan terhadap data yang digunakan dan model yang dibangun agar sesuai dengan kebutuhan perusahaan. Aturan-aturan tersebut yaitu data yang didapatkan dari perusahaan berisi 10 *sample* id *network interface*. Dari 10 sample id tersebut dapat digunakan beberapa id. Id yang akan digunakan tersebut harus memiliki data yang lengkap setiap bulannya agar dapat dianalisis untuk membuat model. Kemudian dari data yang dapat digunakan berdasarkan id, deteksi anomali dilakukan berdasarkan trend normal data *inbound* dan *outbound*.

D. Fase Data Understanding

Berdasarkan penjelasan business understanding pada sub bab sebelumnya, diperlukan data untuk digunakan pada proses penyelesaian masalah yaitu deteksi anomali. Data yang digunakan untuk penelitian ini merupakan *sample* dari data *traffic network element* yang diambil dari MRTG.

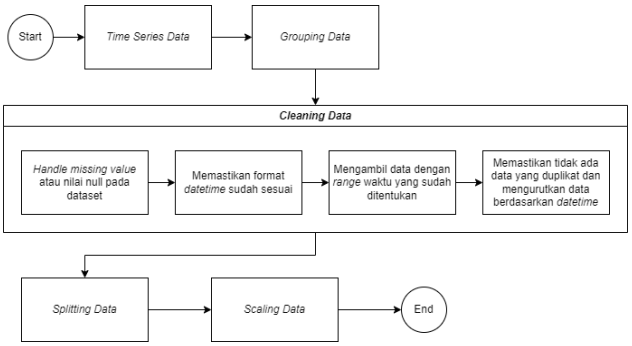
Terdapat *Sample* data memiliki sepuluh id yang terdiri dari id 1, 2, 7, 12, 13, 14, 19, 23, 28, dan 70. Dari sepuluh id tersebut peneliti akan menggunakan delapan id yaitu id 2, 7, 12, 13, 19, 23, 28, dan 70 yang dimana merupakan id yang memiliki data lengkap setiap bulannya. Sedangkan id 2 dan 14 memiliki data *inbound* dan *outbound* yang bernilai nol (0) padahal seharusnya data *inbound* dan *outbound* tidak ada data yang bernilai nol (0). Sample data yang diberikan memiliki rentang waktu Oktober 2021 hingga November 2022. Namun pada data tahun 2021 yaitu data dengan waktu Oktober, November, dan Desember tidak memiliki data yang lengkap setiap bulannya sehingga proses pembuatan model dilakukan dengan menggunakan dt rentang waktu dari 1 Januari 2022 hingga 23 November 2022 yang merupakan tanggal terakhir dari sample data. Data *inbound* dan *outbound* memiliki nilai dengan satuan bit sehingga nilai yang tampil memiliki angka yang tinggi.

Data *sample* awal terdiri dari 708.904 baris data. Data tersebut merupakan data yang diambil dari rentang tanggal 25 Oktober 2021 hingga 23 November 2022. Data *sample* MRTG terdiri dari 10 id yang dimana merupakan id *port* layanan. Id tersebut terdiri dari id 1, 2, 7, 12, 13, 14, 19, 23, 28, dan 70. Namun tidak semua id dapat digunakan dalam penelitian ini dikarenakan terdapat data yang tidak layak untuk dianalisis. Terdapat beberapa id yang memiliki nilai *inbound* dan *outbound* yang bernilai nol (0) yaitu id 1 dan 14 yang memiliki nilai *inbound* dan *outbound* hampir bernilai nol semua. Menurut informasi yang peneliti dapatkan dari hasil wawancara dengan pihak perusahaan, nilai data *inbound* dan *outbound* seharusnya selalu memiliki nilai atau bernilai null. Dikarenakan id 1 dan 14 tidak layak untuk dianalisis berdasarkan informasi yang didapat, maka peneliti tidak akan menggunakan id 1 dan 14.

Setiap id memiliki nilai *inbound* dan *outbound* yang memiliki perbedaan signifikan dikarenakan setiap id memiliki nilai kapasitas yang berbeda-beda. Untuk penelitian deteksi anomali hanya memerlukan data *inbound* dan *outbound* setiap id saja untuk proses deteksi anomali sehingga peneliti tidak menggunakan data kapasitas dari setiap id. Peneliti akan menggunakan data dengan rentang waktu 1 Januari 2022 hingga 23 November 2022 dikarenakan data sebelum 1 Januari 2022 atau data pada tahun 2021 memiliki data berisi *datetime* yang tidak lengkap.

E. Fase Data Preparation

Setelah melakukan *data understanding* pada fase sebelumnya, fase selanjutnya yaitu *data preparation* atau data *pre-processing*. Untuk tahap *data preparation* peneliti akan melakukan beberapa proses yaitu *grouping data, cleaning data, splitting data,* dan *scaling data.* Dapat dilihat pada Gambar III.3 yang merupakan rancangan *data preparation* untuk penelitian ini. Beberapa tahapan yang akan dilakukan dalam *data preparation*. Pertama yaitu *data time series* akan dimasukkan sebagai data yang digunakan untuk tahap *modelling* selanjutnya. Tahap selanjutnya yaitu *grouping data* yang dimana akan membagi data berdasarkan id. Hal tersebut dilakukan karena proses *modelling* untuk deteksi anomali dilakukan per id. Tahap selanjutnya merupakan *cleaning data.* Proses pertama yaitu *handle missing value* atau nilai null dalam dataset. *Handle missing value* dilakukan dengan menggunakan *method interpolate* untuk mengisi nilai null pada data secara linear. Dengan menggunkan *method interpolate* data yang memiliki nilai null akan diisi dengan nilai hasil tebakan berdasarkan nilai yang ada dalam data. Setelah melakukan *handle missing value*, kemudian memastikan kembali apakah sudah tidak ada nilai null pada data dengan melakukan pengecekan nilai null. Memastikan data kolom dt memiliki tipe data datetime dengan format %Y-%m-%d %H:%M:%S sebagai penyesuaian dengan *raw data*. Setelah itu menghapus data dengan rentang waktu dibawah tanggal 1 Januari 2022 karena data yang akan digunakan oleh peneliti merupakan data dengan rentang tanggal 1 Januari hingga akhir yaitu 23 November 2022. Dikarenakan pada *raw data* terdapat data kolom dt yang duplikat padahal seharusnya tidak ada data duplikat, maka peneliti melakukan penghapusan data pada kolom dt yang duplikat dengan menggunakan fungsi *drop_duplicates*. Tahap selanjutnya yang dilakukan yaitu mengurutkan data yang akan digunakan berdasarkan nilai kolom dt yaitu dengan menggunakan urutan dari terlama ke terbaru yakni dari yang memiliki tanggal 1 Januari 2022 hingga 23 November 2022. Sebelum dilakukan tahap *modelling*, akan dilakukan *splitting data* untuk membagi menjadi dua yaitu data latih dan data *testing*. Perbandingan pembagian data latih dan data *testing* yang akan digunakan adalah 70:30. Untuk data latih akan digunakan 70% data dan untuk data *testing* 30% data. Lalu tahap terakhir adalah *scaling data* yang dilakukan untuk menormalisasi data yang digunakan untuk *modelling*.



GAMBAR *III.3*
Rancangan Data Preparation

F. Fase Modelling

Pada fase ini akan dilakukan *modelling* untuk mendeteksi anomali. Penelitian ini menggunakan algoritma *Long Short-Term Memory* (LSTM) untuk membuat model deteksi anomali. Terdapat beberapa tahap yang dilakukan dalam fase *modeling*. Tahap tersebut dimulai dengan *input sample data* yang telah didapatkan dari perusahaan PT XYZ. Kemudian masuk ke tahap *cleaning data*. Tahap *cleaning data* dilakukan berdasarkan hasil perancangan pada sub bab sebelumnya. Setelah data melewati tahap *cleaning data* kemudian data akan dibagi (*split*) menjadi data latih dan data *testing*. Lalu data latih akan digunakan pada tahap pembangunan model LSTM untuk deteksi anomali dan data *testing* akan digunakan untuk tes model LSTM yang telah dihasilkan. Tahap terakhir adalah evaluasi model LSTM yang telah dihasilkan untuk melihat performa model.

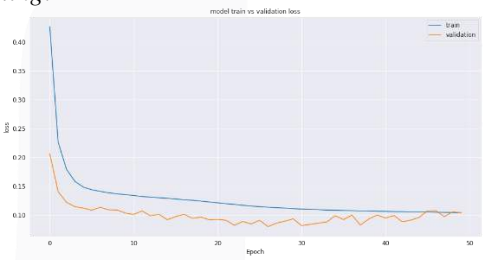
Untuk memahami cara kerja model LSTM yang dibangun diperlukan arsitektur model LSTM. Dalam arsitektur tersebut diperlukan beberapa komponen yang diperlukan dalam proses *modelling*. Komponen tersebut yaitu *input layer, hidden layer, repeat vector layer, dropout layer, dan dense layer*. *Input layer* berperan sebagai masukan untuk data yang akan digunakan. Kemudian *hidden layer* berperan sebagai neuron atau poin yang mengklasifikasikan data dengan menggunakan fungsi aktivasi. Lalu *repeat vector layer* memiliki peran untuk menduplikasi fitur *vector* dari hasil layer sebelumnya untuk mendapat array 2 dimensi untuk proses layer selanjutnya. *Dropout layer* memiliki peran untuk mengurangi jumlah data agar dapat mencegah terjadinya *overfitting*. *Dense layer* memiliki peran sebagai layer keluaran atau output. Desain arsitektur model LSTM ini memiliki 3 *timesteps* yang dimana nilai tersebut berdasarkan data yang digunakan yaitu dengan membandingkan data saat ini dengan data sebelumnya dan data setelahnya yaitu nilai data pada jangka waktu 1 jam yang dimana terdapat 12 data yang telah dikelompokan per 12 data. Deteksi anomali dilakukan berdasarkan membandingkan nilai rata-rata dari kelompok data sekarang dengan rata-rata kelompok data sebelumnya dan rata-rata kelompok data sesudahnya.

IV. HASIL DAN PEMBAHASAN

Fase selanjutnya evaluasi merupakan fase terakhir yang dilakukan dalam penelitian ini. Pada fase ini terdapat dua proses yang dilakukan yaitu proses pengujian (*testing*) dan proses evaluasi. Proses pengujian dilakukan untuk menguji kerja model LSTM yang dibangun untuk mendeteksi anomali pada data *inbound* dan *outbound*. Sedangkan proses evaluasi dilakukan untuk mengevaluasi model LSTM yang sudah dibangun apakah metode yang digunakan sudah tepat dan memberikan hasil terbaik dalam mendeteksi anomali pada data lalu lintas jaringan internal *inbound* dan *outbound*. Pada proses pengujian (*testing*) akan dilakukan pengujian pada model deteksi anomali yang telah dibuat pada proses sebelumnya. Pengujian dilakukan pada setiap id yaitu 2, 7, 12, 13, 14, 23, 28, dan 70. Pengujian ini akan dilakukan berdasarkan jumlah neuron, *batch size*, dan jumlah iterasi (*epoch*) pada data *inboud* dan *outbound* yang berfokus pada id 23.

Pengujian data inbound dilakukan pada tren data, nilai anomali, dan nilai anomali terhadap data id 23. Pada nilai anomali ditentukan nilai batas (*threshold*) dari nilai anomali yang ditentukan. Nilai batas yang digunakan dalam penelitian ini terdapat dua nilai batas. Nilai batas pertama ditentukan berdasarkan rata-rata dari perbandingan selisih nilai data sekarang dengan data sebelumnya. Sedangkan nilai batas kedua ditentukan berdasarkan rata-rata perbandingan selisih nilai data sekarang dengan data setelahnya.

Pada Gambar IV.1 menunjukan grafik tren data inbound loss dari data latih (training) dan data testing terhadap jumlah iterasi (*epoch*) yang dilakukan yaitu sejumlah 50 iterasi. Grafik menunjukan nilai loss data latih dan data *testing* pada setiap iterasi memiliki nilai yang bersinggungan sehingga data tidak *overfitting* maupun *underfitting*.



GAMBAR IV.1
Grafik Tren Data Inbound Id 23

Pada Gambar IV.2 menampilkan data anomali yang terdapat pada data *inbound* id 23. Terdapat 93 data yang didapatkan sebagai anomali. Data anomali terdapat pada data dengan tanggal 10 Juli 2022, 17 Juli 2022, 31 Juli 2022, 7 Agustus 2022, 14 Agustus 2022, 21 Agustus 2022, dan 10 November 2022.

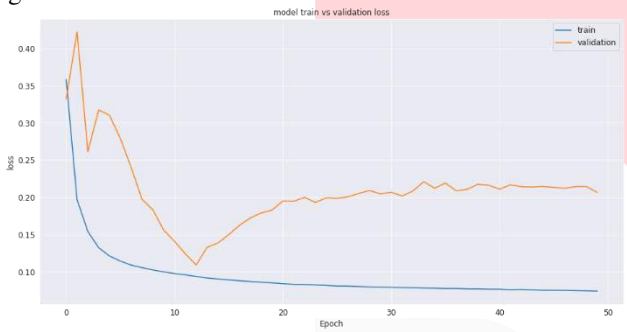
dt	Inbound Sc	Anomali
2022-07-10 02:35:00	-0.758354	True
2022-07-10 02:40:00	-0.804354	True
2022-07-10 02:45:00	-0.851992	True
2022-07-10 02:50:00	-0.898498	True
2022-07-10 02:55:00	-0.943977	True
...	...	...
2022-11-10 20:10:00	1.322309	True
2022-11-10 20:15:00	1.355375	True
2022-11-10 20:20:00	1.393133	True
2022-11-10 20:25:00	1.426525	True
2022-11-10 20:30:00	1.465411	True

93 rows × 2 columns

GAMBAR IIV.2
Data Anomali Inbound Id 23

Pengujian data *outbound* dilakukan pada tren data, nilai anomali, dan nilai anomali terhadap data id 23. Pada nilai anomali ditentukan nilai batas (*threshold*) dari nilai anomali yang ditentukan. Nilai batas yang digunakan dalam penelitian ini terdapat dua nilai batas. Nilai batas pertama ditentukan berdasarkan rata-rata dari perbandingan selisih nilai data sekarang dengan data sebelumnya. Sedangkan nilai batas kedua ditentukan berdasarkan rata-rata perbandingan selisih nilai data sekarang dengan data setelahnya.

Pada Gambar IV.3 menunjukan grafik tren data *outbound loss* dari data latih (*training*) dan data testing terhadap jumlah iterasi (*epoch*) yang dilakukan yaitu sejumlah 50 iterasi. Grafik menunjukan nilai loss data latih dan data testing pada setiap iterasi memiliki nilai yang konsisten dan cukup baik memiliki selisih loss yang tidak signifikan.



GAMBAR IIV.3
Grafik Tren Data Outbound Loss Id 23

Pada Gambar IV.4 menampilkan data anomali yang terdapat pada data *outbound* id 23. Terdapat 55 data yang didapatkan sebagai anomali. Data anomali terdapat pada data dengan tanggal 10 Juli 2022, 17 Juli 2022, 31 Juli 2022, dan 14 Agustus 2022.

Pada proses evaluasi dilakukan untuk menganalisis hasil model LSTM pada setiap id. Proses evaluasi ini dilakukan dengan tujuan untuk membandingkan hasil *fit model* mana yang terbaik berdasarkan parameter yang diterapkan. Selain itu, evaluasi juga digunakan untuk memastikan model sudah sesuai dengan studi kasus yang sedang diteliti.

dt	Inbound Sc	Anomali
2022-07-10 02:20:00	0.769214	True
2022-07-10 02:25:00	0.705856	True
2022-07-10 02:30:00	0.639716	True
2022-07-10 02:35:00	0.712278	True
2022-07-10 02:40:00	0.648409	True
2022-07-10 02:45:00	0.583071	True
2022-07-10 02:50:00	0.518779	True
2022-07-10 02:55:00	0.455349	True
2022-07-10 03:00:00	0.392079	True
2022-07-10 03:05:00	0.330776	True
2022-07-10 03:10:00	0.269460	True
2022-07-10 03:15:00	0.209210	True
2022-07-10 03:20:00	0.150495	True
2022-07-10 03:25:00	0.093873	True
2022-07-10 03:30:00	0.038722	True
2022-07-17 02:55:00	0.126725	True

GAMBAR IIV.4
Data Anomali Outbound Id 23

Pada model LSTM di data inbound id 23 menghasilkan nilai evaluasi MAE secara keseluruhan yaitu 0.06722991287041966. Berdasarkan nilai tersebut, model LSTM yang dihasilkan baik dan cukup akurat. Dan Pada model LSTM di data *outbound* id 23 menghasilkan nilai evaluasi MAE secara keseluruhan yaitu 0.08836275776405246. Berdasarkan nilai tersebut, model LSTM yang dihasilkan baik dan cukup akurat.

V. KESIMPULAN

A. Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, rancangan deteksi anomali pada lalu lintas jaringan internal inbound dan outbound menggunakan model dengan algoritma *Long Short-Term Memory* (LSTM) dapat menghasilkan model yang mendeteksi data anomali dengan baik pada setiap id berdasarkan nilai MAE yang didapatkan. Dari hasil deteksi anomali yang didapatkan dalam penelitian ini memberikan hasil yang efektif dalam mendeteksi anomali dari data lalu lintas jaringan internal inbound dan outbound. Beberapa proses yang dilakukan dalam mendeteksi anomali yaitu *pre-processing data, modelling* untuk pelatihan dan *testing*, evaluasi model, dan deteksi anomali. Proses yang dilakukan dengan berurutan memberikan nilai MAE tidak melebihi 0.08.

REFERENSI

[1] M. Said Elsayed, N. A. Le-Khac, S. Dev, and A. D. Jurcut, “Network Anomaly Detection Using LSTM Based Autoencoder,” in *Q2SWinet 2020 - Proceedings of the 16th ACM Symposium on QoS and Security for Wireless and Mobile Networks*, Nov. 2020, pp. 37–45. doi: 10.1145/3416013.3426457.

[2] F. Martinez-Plumed *et al.*, “CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories,” *IEEE Trans Knowl Data Eng*, vol. 33, no. 8, pp. 3048–3061, Aug. 2021, doi: 10.1109/TKDE.2019.2962680.

[3] A. Gulli, A. Kapoor, S. Pal, O’Reilly for Higher Education (Firm), and an O. M. Company. Safari, *Deep Learning with TensorFlow 2 and Keras - Second Edition*.

[4] H. D. Nguyen, K. P. Tran, S. Thomassey, and M. Hamad, “Forecasting and Anomaly Detection approaches using LSTM and LSTM Autoencoder techniques with the applications in supply chain management,” *Int J Inf Manage*, vol. 57, Apr. 2021, doi: 10.1016/j.ijinfomgt.2020.102282.