BAB III
METODE PENELITIAN

Pada penyususnan tesis ini diperlukan adanya data yang akurat dan dipercaya dalam menyelesaikan suatu permasalahan yang akan dicari solusi yang tepat. Maka di perlukan suatu metode yang sesuai dengan permasalahan yang sesuai dengan topik penelitian. Di harapkan dengan menggunakan metode dapat memberikan hasil yang sesuai dengan yang di harapkan dan dapat di pertanggung jawabkan secara ilmiah, sehingga data yang di dapat adalah data yang *objektif, valid dan rilabel.* Metode penelitian penelitian itu sendiri merupakan suatu prosedur atau cara untuk mengetahui sesuatu dengan langkah-langkah sistematis untuk mendapatkan fakta-fakta atau prinsip-prinsip baru yang bertujuan untuk mendapatkan pengertian atau hal-hal baru dan menaikan tingkat ilmu serta teknologi.

3.1.Sumber Data

Pengambilan dataset DDoS ini menggunakan data yang di terbitkan oleh dataset jaringan DDoS evaluation dataset (CIC-DDoS2019) yang di tulis oleh canadian institutfor cyber security. Rekaman data mentah termasuk lalu lintas jaringan (Pcaps) dan log peristiwa (Windows dan Ubuntu event Logs) per mesin. Dalam proses ekstraksi fitur dari data mentah ini menggunakan CICFlowMeter-V3 dan mengekstrak 88 fitur/atribut dan menyimpannya sebagai file CSV per mesin[48].

Gambar 3. 1 Jumlah Record Data dan Atribut

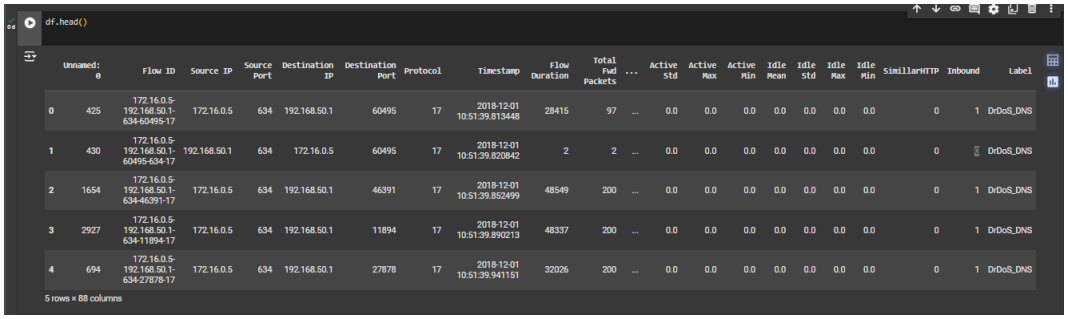
Jumlah record dalam dataset ini yaitu 9.807.780 *record* dengan distribusi recordnya adalah sebagai berikut:

Tabel 3. 1 Data Distribusi Jumlah data DDoS

No	Jenis Serangan	Record
1	DrDoS_SSDP	1048271
2	Syn	1048228
3	DrDoS_LDAP	1047795

No	Jenis Serangan	Record
4	DrDoS_NetBIOS	1047706
5	DrDoS_MSSQL	1047561
6	DrDoS_UDP	1047441
7	TFTP	1047402
8	DrDoS_DNS	1046567
9	DrDoS_NTP	1035009
10	UDP-lag	366461
11	WebDDoS	439
12	BENIGN	24900

Log yang ada pada *dataset* DDoS seperti yang tersaji pada Gambar 3.2 berikut.



df.head()

	Unnamed: 0	Flow ID	Source IP	Source Port	Destination IP	Destination Port	Protocol	Timestamp	Flow Duration	Total Fwd Packets	...	Active Std	Active Max	Active Min	Idle Mean	Idle Std	Idle Max	Idle Min	SimillarHTTP	Inbound	Label
0	425	172.16.0.5-192.168.50.1-634-60495-17	172.16.0.5	634	192.168.50.1	60495	17	2018-12-01 10:51:39.813448	28415	97	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0	1	DrDoS_DNS
1	430	172.16.0.5-192.168.50.1-60495-634-17	192.168.50.1	634	172.16.0.5	60495	17	2018-12-01 10:51:39.820842	2	2	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0	0	DrDoS_DNS
2	1654	172.16.0.5-192.168.50.1-634-46391-17	172.16.0.5	634	192.168.50.1	46391	17	2018-12-01 10:51:39.852499	48549	200	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0	1	DrDoS_DNS
3	2927	172.16.0.5-192.168.50.1-634-11894-17	172.16.0.5	634	192.168.50.1	11894	17	2018-12-01 10:51:39.890213	48337	200	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0	1	DrDoS_DNS
4	694	172.16.0.5-192.168.50.1-634-27878-17	172.16.0.5	634	192.168.50.1	27878	17	2018-12-01 10:51:39.941151	32026	200	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0	1	DrDoS_DNS

5 rows × 88 columns

Gambar 3. 2 Traffics Collection Dataset

Berdasarkaan data log di atas terdapat 88 fitur/atribut yang di buat. Berikut akan dijelaskan pada tabel 3.2 untuk detail fitur/atribut yang ada pada *Dataset* DDoS 2019.

Tabel 3. 2 Fitur / Atribur Dataset CIC-DDoS2019[49]

No	**Feature Name**	**Description**	**No**	**Feature Name**	**Description**
1	Unnamed: 0		45	Bwd Packets/s	Number of backward packets per second
2	Flow ID	Flow Identity	46	Min Packet Length	Minimum length of a flow
3	Source IP	Source IP ddress	47	Max Packet Length	Maximum length of a flow
4	Source Port	Source Port	48	Packet Length Mean	Mean length of a flow

No	Feature Name	Description	No	Feature Name	Description
5	Destination IP	Destination IP Address	49	Packet Length Std	Standard deviation length of a flow
6	Destination Port	Destination Port	50	Packet Length Variance	Minimum inter arrival time
7	Protocol	Protocol	51	FIN Flag Count	Number of packets with FIN
8	Timestamp	Date format that is distributed on nix-based servers	52	SYN Flag Count	Number of packets with SYN
9	Flow Duration	Flow Duration	53	RST Flag Count	Number of packets with RST
10	Total Fwd Packets	Total Forward Packets	54	PSH Flag Count	Number of packets with PSH
11	Total Backward Packets	Total Backward Packets	55	ACK Flag Count	Number of packets with ACK
12	Total Length of Fwd Packets	Total Length Forward Packets	56	URG Flag Count	Number of packets with URG
13	Total Length of Bwd Packets	Total Length Backward Packets	57	CWE Flag Count	Number of packets with CWE
14	Fwd Packet Length Max	Forward Packet Length Max	58	ECE Flag Count	Number of packets with ECE
15	Fwd Packet Length Min	Forward Packet Length Min	59	Down/Up Ratio	Download and upload ratio
16	Fwd Packet Length Mean	Forward Packet Length Mean	60	Average Packet Size	Average size of packet
17	Fwd Packet Length Std	Forward Packet Length Standard	61	Avg Fwd Segment Size	Average size bserved in the forward direction
18	Bwd Packet Length Max	Backward Packet Length Max	62	Avg Bwd Segment Size	Average size observed in the backward direction
19	Bwd Packet Length Min	Backward Packet Length Min	63	Fwd Header Length.1	Forward Header Length
20	Bwd Packet Length Mean	Backward Packet Length Mean	64	Fwd Avg Bytes/Bulk	Average number of bytes bulk rate in the forward direction
21	Bwd Packet Length Std	Backward Packet Length Standard	65	Fwd Avg Packets/Bulk	Average number of packets bulk rate in the forwarddirection

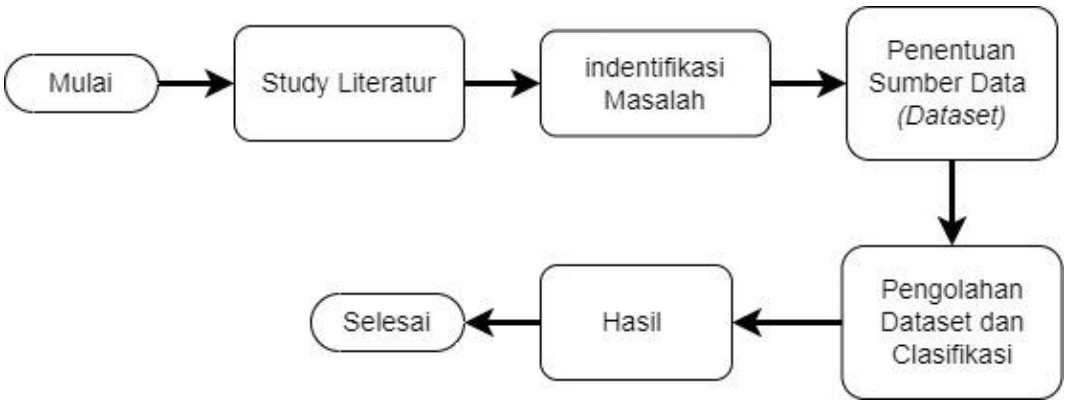
No	Feature Name	Description	No	Feature Name	Description
22	Flow Bytes/s	flow byte rate that is number of packets transferred per second	66	Fwd Avg Bulk Rate	Average number of bulk rate in the forward direction
23	Flow Packets/s	Flow packets rate that is number of packets transferred per second	67	Bwd Avg Bytes/Bulk	Average number of bytes bulk rate in the backward direction
24	Flow IAT Mean	Mean time between two flows	68	Bwd Avg Packets/Bulk	Average number of packets bulk rate in the backward direction
25	Flow IAT Std	Standard time between two flows	69	Bwd Avg Bulk Rate	Average number of bulk rate in the backward direction
26	Flow IAT Max	Max time between two flows	70	Subflow Fwd Packets	The average number of packets in a sub flow in the forward direction
27	Flow IAT Min	Min time between two flows	71	Subflow Fwd Bytes	The average number of bytes in a sub flow in the forward direction
28	Fwd IAT Total	Total time between two packets sent in the forward direction	72	Subflow Bwd Packets	The average number of packets in a sub flow in the backward direction
29	Fwd IAT Mean	Mean time between two packets sent in the forward direction	73	Subflow Bwd Bytes	The average number of bytes in a sub flow in the backward direction
30	Fwd IAT Std	Standard deviation time between two packets sent in the forward direction	74	Init_Win_bytes_forward	Number of bytes sent in initial window in the forward direction
31	Fwd IAT Max	Maximum time between two packets sent in the forward direction	75	Init_Win_bytes_backward	Number of bytes sent in initial window in the backward Zirection

No	Feature Name	Description	No	Feature Name	Description
32	Fwd IAT Min	Minimum time between two packets sent in the forward direction	76	act_data_pkt _fwd	Number of packets with at least 1 byte of TCP data payload in the forward direction
33	Bwd IAT Total	Maximum time between two packets sent in the forward direction	77	min_seg_size _forward	Minimum segment size observed in the forward direction
34	Bwd IAT Mean	Mean time Between two packets sent in the backward direction	78	Active Mean	Mean time a flow was active before becoming idle
35	Bwd IAT Std	Standard deviation time between two packets sent in the backward direction	79	Active Std	Standard deviation time a flow was active before becoming idle
36	Bwd IAT Max	Maximum time between two packets sent in the backward direction	80	Active Max	Maximum time a flow was active before becoming idle
37	Bwd IAT Min	Minimum time between two packets sent in the backward direction	81	Active Min	Minimum time a flow was active before becoming idle
38	Fwd PSH Flags	Numberof times the PSH flag was set in packets travelling in the forward direction	82	Idle Mean	Mean time a flow was idle before becoming active
39	Bwd PSH Flags	Number of times the PSH flag was set in packets travelling in the backward	83	Idle Std	Standard deviation time a flow was idle before becoming active
40	Fwd URG Flags	Number of times the URG flag was set in packets travelling in the forward direction (0 for UDP)	84	Idle Max	Maximum time a flow was idle before becoming active

No	Feature Name	Description	No	Feature Name	Description
41	Bwd URG Flags	Number of times the URG flag was set in packets travelling in the backward direction (0 for UDP)	85	Idle Min	Minimum time a flow was idle before becoming active
42	Fwd Header Length	Total bytes used for headers in the Forward direction	86	SimillarHTTP	HTTP Simillarity
43	Bwd Header Length	Total bytes used for headers in the Backward direction	87	Inbound	Inbound Traffic
44	Fwd Packets/s	Number of forward packets per Second	88	Label	Label Attack

3.2.Tahapan penelitian

Langkah-langkah yang di di lakukan ada penelitian ini di gambarkan mengunakan diagram alir. Langkah pertama dimulai dari studi literatur terhadap topik yang dipilih berdasarkan permasalahan-permasalahan yang ada, kemudian di lanjutkan dengan pembelajaran terkait penelitian yang sudah di lakukan, setelah itu melakukan pengumpulan datasetdan dilanjutkan dengan pelatihan dan pengujian data yang telah di kumpulkan, hasil kemudian di analisis dan di tarik kesimpulan.



Gambar 3. 3 Tahapan Penelitian

3.2.1 Studi Literatur

Tahapan yang pertama kali dilakukan yaitu studi lieratur atau studi awal yang merupakan tahap yang pertama kali dilakukan dalam penelitian. Pada tahap ini, peneliti melakukan observasi atau pemahaman penelitian yang meliputi tujuan dan permasalahan yang di angkat secara keseluruhan, menerjemahkan tujuan dan batasan ke dalam perumusan definisi masalah deteksi DDoS Attact. Dalam studi literatur ini juga dilakukan pencarian terhadap sumber-sumber teori yang relevan dengan topik penelitian dari berbagai sumber seperti buku, penelitian terkait yang ada dalam suatu jurnal, internet, dan lain sebagainya yang mendukung proses penelitian dengan tujuan untuk memperkuat permasalahan.

3.2.2 Identifikasi Masalah

Identifikasi masalah merupakan bagian dari proses penelitian yang dapat dipahami sebagai upaya mendefinisikan problem serta membuat definisi tersebut menjadi lebih terukur atau measurable sebagai suatu langkah awal penelitian. Pada langkah ini permasalahan keamanan cyber menjadi isu yang cukup populer yang terjadi pada dunia teknologi informasi dan jaringan, salah satunya serangan DDoS attact yang menghabiskan resource yang cukup besar sehingga menyebabkan server down.

3.2.3 Penentuan Sumber Data

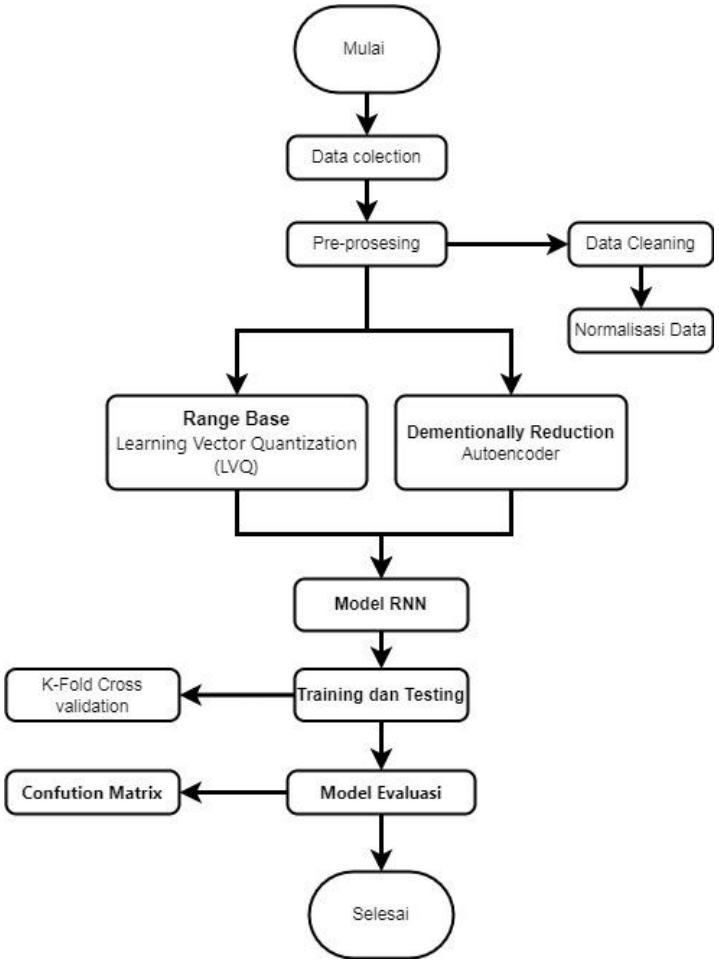
Data yang digunakan dalam penelitian ini adalah data publik yaitu data tentang DDoS Attack. Pengumpulan data menggunakan analisis data eksplorasi untuk membiasakan diri dengan data dan menemukan wawasan awal dan mengevaluasi kualitas data. Data tersebut merupakan data terbaru yang di keluarkan oleh Canadian Institute for Cybersecurity yang di keluarkan pada tahun 2019.

3.2.4 Pengolahan Dataset dan Kasifikasi

Setelah dataset didapat maka akan dilakukan proses pengolahan data mulai dari preprosesing yang di gunakan untuk memperbaikai data yang kurang sesuai, pemilihan fitur memiliki fungi untuk menentukan komponen paling peting saja yang di ambil tanpa mengurangi data penting yang ada dalam dataset, dan

pemisahkan data ke dalam set pelatihan dan pengujian untuk mengevaluasi model penggalian data. Saat memisahkan himpunan data ke dalam set pelatihan dan set pengujian, sebagian besar data digunakan untuk pelatihan, dan bagian data yang lebih kecil digunakan untuk pengujian. Dalam Mechine learning ini sampel data secara acak untuk membantu memastikan bahwa set pengujian dan pelatihan serupa. Dengan menggunakan data serupa untuk pelatihan dan pengujian, dapat meminimalkan efek perbedaan data dan lebih memahami karakteristik model.

Pada penelitian ini pengolahan dataset dan clasifikasi serangan menggunakan pemprograman phyton sebagai alat data mining untuk mengevaluasi dan menentukan serangan DDoS attack. Analisaa dan proses kalsifikasi menggunakan dataset dari CICDDoD2019 dengan menggunakan K-fold cross validation dengan algoritma Recurent Neural Network berikut tahapan analisis dan kalasifikasi serangan DDoS.



Gambar 3. 4 Diagram pengolahan data dan clasifikasi

3.2.4.1 Data Colection

Pada tahap pengumpulan data dengan tujuan untuk mendapatkan data yang akan diolah pada penelitian. Data yang digunakan dalam penelitian ini adalah data public yaitu data CIC-DDoS2019. Data tersebut merupakan data terbaru yang di keluarkan oleh Canadian Institute for Cybersecurity, dengan rincian data dijabarkan dalam dataset.

3.2.4.2 Pre-Prosesing

Dalam tahapan ini data di persiapkan dengan matang, proses ini mengubah data mentah ke dalam bentuk yang lebih mudah dipahami. Proses ini diperlukan untuk memperbaiki kesalahan pada data mentah yang seringkali tidak lengkap dan memiliki format yang tidak teratur. Preprocessing data penting dilakukan untuk memastikan kualitas data dan mempermudah proses analisis data.

1. **Data cleaning**

   Data set yang dikumpulkan pasti memiliki missing value atau noise. Hal ini karena proses pengumpulan datanya tidak sempurna sehingga ada banyak bagian yang tidak relevan dan hilang. pada tahapan ini data cleaning merupakan proses mendeteksi, mengoreksi, atau menghapus data yang tidak akurat, tidak lengkap, atau tidak relevan dalam suatu dataset untuk meningkatkan kualitas analisis. Dalam data cleaning terdapat beberapa langkah yaitu missing value, duplikat data, dan penanganan outlier.

2. **Normalisasi**

   Normalisasi data adalah proses transformasi data ke dalam skala yang lebih seragam untuk meningkatkan akurasi analisis dan kinerja algoritma pembelajaran mesin. Normalisasi membantu mengatasi masalah skala yang berbeda dalam dataset sehingga setiap fitur memiliki kontribusi yang seimbang.

3.2.4.3 Seleksi Fitur

Setelah tahap pra-pemrosesan data selesai dan jumlah fitur target telah ditetapkan, langkah krusial berikutnya adalah seleksi fitur (*feature selection*). Proses ini

merupakan tahapan fundamental dalam pembangunan model machine learning, di mana dilakukan pemilihan sub-kumpulan (*subset*) atribut yang paling relevan dan informatif dari keseluruhan dataset.Tujuan utamanya adalah untuk meningkatkan performa model deteksi secara keseluruhan. Dengan mereduksi dimensi data, model dapat menghilangkan fitur-fitur yang bersifat redundan, tidak informatif, atau bahkan *noisy* (mengganggu). Hal ini memungkinkan algoritma klasifikasi untuk lebih fokus pada sinyal atau pola yang benar-benar signifikan dalam membedakan lalu lintas normal dan serangan. Sebagai hasilnya, proses seleksi fitur tidak hanya berkontribusi pada peningkatan akurasi, tetapi juga secara signifikan mempercepat waktu yang dibutuhkan untuk proses pelatihan dan inferensi, serta mengurangi risiko overfitting.

A. Learning Vector Quatizition

LVQ bekerja dengan mempelajari prototypes (vektor representatif) untuk setiap kelas dalam data. Selama proses pelatihan, LVQ memodifikasi prototypes berdasarkan jarak antara data dan prototypes. Fitur-fitur yang paling berpengaruh dalam menentukan jarak ini dapat dianggap sebagai fitur yang paling penting. Dengan menganalisis kontribusi setiap fitur terhadap pembentukan prototypes, kita dapat mengidentifikasi fitur-fitur yang paling relevan untuk tugas klasifikasi. Fitur-fitur ini kemudian dapat dipilih sebagai subset fitur yang digunakan dalam model machine learning.

Tabel 3. 3 Parameter LVQ

Parameter	Nilai
epoch awal	0
maxepoch	100
Pengurangan pembelalajaran	0.1
Learning rate	0.00001
Minimum learning rate	0.000001
Prototype per kelas	5
Fitur yang dipilih	16

B. Autoencoder

Pemilihan fitur dengan *autoencode*r bertujuan untuk mengurangi data tanpa menghapus informasi penting serta memperkecil ukuran dataset, sehingga dapat meningkatkan kinerja deteksi[50]. Jaringan saraf buatan ini dibuat untuk mempelajari serta merekonstruksi representasi ciri. Arsitekturnya terdiri dari dua *elemen simetris*: pengkode dan pengurai. Encoder bertanggung jawab untuk mengekstrak fitur dari dataset, sedangkan decoder bertugas merekonstruksi data berdasarkan fitur-fitur itu. Apabila jumlah neuron di lapisan tersembunyi lebih rendah daripada jumlah neuron di lapisan input, model akan melakukan kompresi data. Selama proses pelatihan, model pembelajaran menciptakan representasi berdimensi lebih rendah dari data asli dengan mempertahankan informasi sebesar mungkin. Dalam percobaan ini, diterapkan arsitektur AE yang ringkas karena sesuai dengan ciri-ciri data[38].

Tabel 3. 4 Parameter Autoencoder [50]

Parameter	Nilai
Node input	85
Node output	85
Node hiden layer 1	64
Node hiden layer 2	32
Node hiden layer 3	16
Activation function of hidden layer	Relu
Activation function of Ouput layer	Sigmoid and Softmax
Learning rate	0.00001
Fungsi Loss	Mean Squared Error (MSE)
Fungsi optimasi Optimization	Adam
epoch	100

C. Gabungan fiture selection (LVQ-Autoencoder)

Seleksi fitur ini merupakan kombinasi dari dua seleksi yang telah dilakukan sebelumnya. Pendekatan ini mengintegrasikan dua metode yang berbeda, yaitu *Learning Vector Quantization* (LVQ) dan *Autoencoder*. Tujuan pendekatan ini

adalah untuk mendapatkan subset fitur terbaik dari data asli dengan mempertimbangkan dua sudut pandang yang berbeda dalam menilai pentingnya fitur. LVQ memberikan bobot sesuai kontribusi fitur dalam membedakan kelas melalui proses pembelajaran terawasi dengan menambahkan bobot pada tiap fitur, sehingga dapat terbentuk fitur yang paling dominan dan memiliki skor kepentingan yang baik. Autoencoder menetapkan bobot berdasarkan kemampuan fitur dalam merekonstruksi data melalui metode pembelajaran tanpa pengawasan. Model autoencoder adalah jaringan saraf yang memiliki input dan output yang identik, digunakan untuk mengurangi dimensi dari fitur (*Dimensionality Reduction*). Langkah-langkah yang diterapkan dalam seleksi sebagai berikut:

1. Mengambil Skor Pentingnya Fitur dari LVQ

   Setelah pelatihan model LVQ rampung, didapatkan bobot penting fitur yang menunjukkan sumbangan setiap fitur dalam memisahkan kelas. Bobot ini telah dinormalisasi dan jumlahnya sama dengan jumlah fitur yang asli.
2. Mengambil Skor Pentingnya Fitur dari Autoencoder

   Dalam model Autoencoder, nilai fitur ditentukan berdasarkan perannya dalam proses pemulihan data. Outputnya disimpan sebagai array yang mencerminkan sensitivitas masing-masing fitur terhadap hasil Autoencoder.
3. Normalisasi Skor Autoencoder

   Skor relevansi fitur dari Autoencoder dinormalisasi dengan Min-Max Scaling sehingga memiliki rentang nilai antara 0 dan 1. Langkah ini diambil untuk menjamin bahwa skor dari Autoencoder dapat diintegrasikan dengan adil bersama bobot dari LVQ.
4. Penggabungan Skor Pentingnya Fitur

   Skor dari LVQ dan Autoencoder dijumlahkan lalu dirata-rata untuk mendapatkan skor komposit dari setiap fitur. Pendekatan penjumlahan dasar ini diterapkan karena kedua cara dianggap memberikan kontribusi yang sama penting.
5. Mengurutkan Fitur Berdasarkan Skor Gabungan

   Fitur-fitur disusun menurut skor total tersebut secara menurun. Urutan ini menunjukkan fitur yang memiliki kombinasi kontribusi terbaik dari kedua

metode.

6. Pemilihan Top-N Fitur

   Berdasarkan hasil urutan, dipilih 16 fitur terbaik sebagai fitur hasil seleksi gabungan/hybrid.

7. Pembuatan Dataset Baru

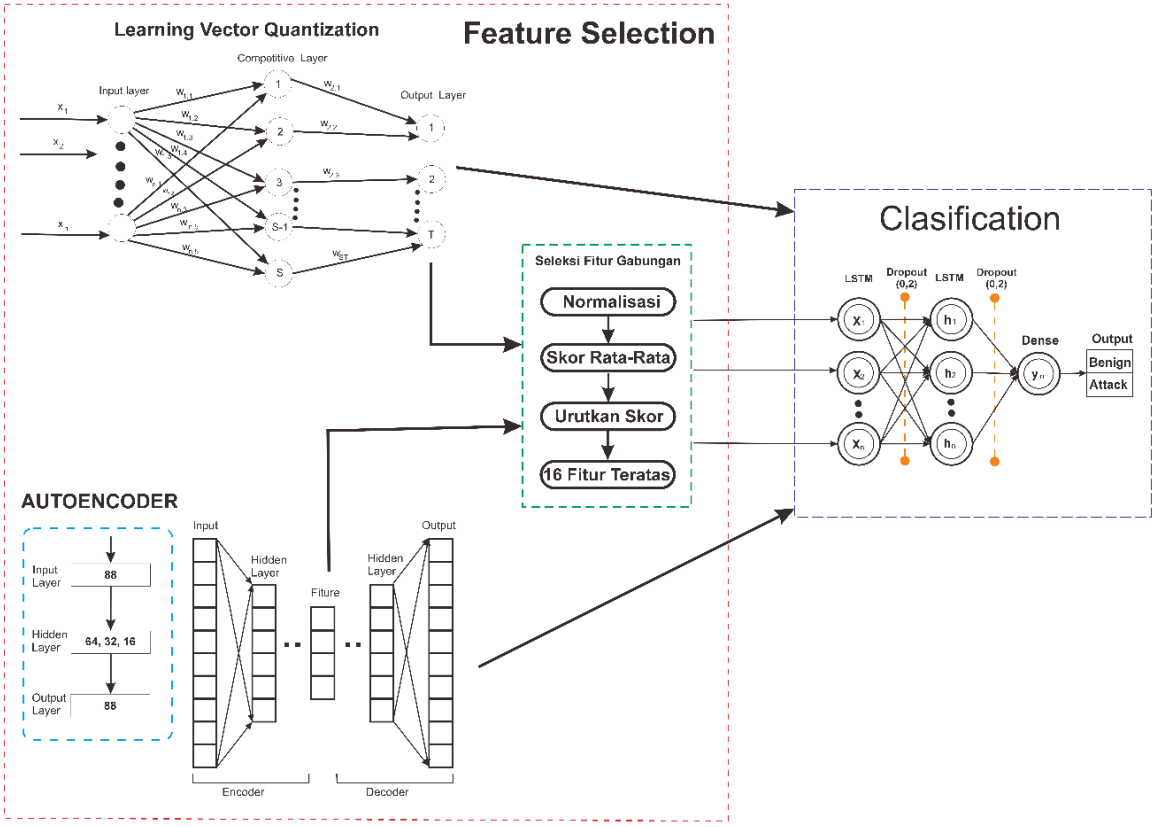
   Dataset baru dibentuk yang hanya mencakup fitur-fitur hasil seleksi hybrid dan label target. Dataset ini selanjutnya disimpan dalam format CSV agar dapat digunakan pada tahap pelatihan model klasifikasi berikutnya.

Pemilihan 16 fitur sebagai jumlah akhir untuk reduksi dimensi dalam penelitian ini didasarkan pada desain arsitektur model, konsistensi untuk perbandingan yang setara, serta keselarasan dengan temuan pada penelitian-penelitian terkait. Berdasarkan asitekturnya model seleksi telah ditetapkan dengan parameter autoencoder berdasarkan penelitian sebelumnya, yang menjelaskan bahwa terdapat 16 fitur esensial yang paling representatif terhadap proses clasifikasi anomali[50]. Selain itu juga reduksi dimensi di lakukan secara signifikan hingga ke jumlah fitur yang spesifik (bukan berdasarkan persentase) sejalan dengan praktik yang telah terbukti efektif dalam literatur deteksi anomali jaringan [23]. Setelah angka 16 ditetapkan dan divalidasi oleh studi eksternal, jumlah ini kemudian diadopsi secara konsisten untuk metode seleksi fitur lainnya (LVQ dan Hybrid). Tahap ini di lakukan untuk memastikan bahwa perbandingan kinerja antarskenario model dilakukan secara adil dan objektif. Dengan menggunakan kuantitas fitur yang sama, perbedaan performa yang teramati dapat lebih valid diatribusikan pada **kualitas dan relevansi** fitur yang dipilih oleh masing-masing algoritma.

3.2.4.4 Modeling Recurenct Neural Network

Recurrent Neural Network (RNN) adalah salah satu jenis *neural network* yang dikhususkan untuk mengolah data berurutan atau time series, di mana urutan dan konteks waktu sangat krusial. RNN mampu "mengingat" data dari langkah-langkah sebelumnya dan memanfaatkannya untuk mempengaruhi hasil saat ini. Dalam arsitektur RNN terdapat 3 komponen yaitu input, lapisan tersembunyi, dan output.

Inputan adalah data dari tahap sebelumnya pada waktu tertentu, hiden layer menyimpan informasi dari tahap sebelumnya, sedangkan output adalah hasil prediksi yang didasarkan pada *hiden layer*. Model ini dimulai dengan lapisan input yang mendapatkan data dalam bentuk urutan fitur *time-series*, di mana setiap *timestep* merepresentasikan informasi lalu lintas jaringan pada waktu tertentu. Selanjutnya, model ini dilengkapi dengan dua lapisan LSTM bertingkat. Lapisan pertama memiliki 64 unit dengan parameter return *sequences = True* untuk menjaga *output* dalam format urutan, sehingga dapat diproses lebih lanjut oleh lapisan selanjutnya. Setelah itu, diterapkan lapisan *dropout* dengan rasio 0.2 untuk mencegah *overfitting* sebelum masuk ke lapisan LSTM kedua yang memiliki 32 unit dengan *return sequences = False*, yang berfungsi untuk menangkap pola fitur yang lebih abstrak. Model selanjutnya berlanjut ke lapisan *Dense* dengan 16 unit dan fungsi aktivasi ReLU yang berfungsi untuk mengubah fitur menjadi format yang lebih sesuai untuk klasifikasi. Akhirnya, lapisan output memanfaatkan 1 unit dengan aktivasi sigmoid untuk menghasilkan peluang klasifikasi biner di antara 0 dan 1 (normal atau DDoS). Model ini dioptimalkan dengan *Adam Optimizer*, yang dipilih karena kemampuannya dalam mempercepat konvergensi serta menyesuaikan parameter secara adaptif. Sebagai fungsi kerugian, digunakan *Binary Crossentropy* karena model melakukan klasifikasi dua kelas.



Gambar 3. 5 Asitektur sistem yang di usulkan

3.2.4.5 Traning dan Testing

1. Penyeimbangan Data (Data Balancing)

Proses ini sangat krusial untuk menangani ketidakseimbangan dalam dataset, di mana satu atau beberapa kelas memiliki jumlah sampel yang jauh lebih banyak atau lebih sedikit dibandingkan kelas yang lain. Ketidakseimbangan data sering muncul dalam isu klasifikasi, termasuk dalam deteksi serangan siber seperti DDoS, di mana jumlah lalu lintas normal biasanya jauh lebih tinggi dibandingkan dengan traffic serangan. Apabila tidak diatasi, ketidakseimbangan data dapat membuat model machine learning cenderung kepada kelas mayoritas, yang berarti mengurangi akurasi deteksi untuk kelas minoritas. Penyeimbangan data dilaksanakan pada data latihan untuk setiap lipatan yang dijalankan.

2. K-Fold Cross Validation

Proses pelatihan untuk menguji kecocokan fitur ini dilaksanakan dengan *K-fold cross validation* guna memanfaatkan seluruh dataset yang telah menjalani seleksi fitur secara optimal. Dalam pelatihan ini, dilakukan uji Validasi silang sepuluh kali lipat yang diterapkan pada setiap lipatan untuk memilih fitur optimal, dan fitur yang terpilih kemudian ditabulasikan. Fitur yang optimal ditentukan pada akhir proses validasi silang 10 kali lipat berdasarkan jumlah kemunculan[27]. Untuk menghindari *overfitting*, diterapkan mekanisme *Early Stopping*, yaitu teknik penghentian lebih awal yang secara otomatis menghentikan proses pelatihan jika kinerja model pada data validasi tidak mengalami kemajuan setelah sejumlah epoch tertentu (*patience*). Ini mendukung jaminan bahwa model mampu menggeneralisasi dengan baik pada data yang baru[51].

Tahapan dalan uji validasi ini sebagai berikut:

a. Dataset dipecah menjadi K subset (folds) yang memiliki ukuran hampir setara.
b. Model dilatih dengan menggunakan K-1 subset sebagai data pelatihan dan diuji pada 1 subset sebagai data pengujian.
c. Proses ini diulang K kali, dengan setiap lipatan digunakan sekali sebagai data uji.
d. Rata-rata hasil evaluasi dari setiap iterasi digunakan untuk memperoleh estimasi performa model yang lebih tepat..

3.2.4.6 Evaluation matrix

Dalam studi ini, kriteria penilaian diterapkan pada pengujian dataset CIC-DDoS2019. Evaluasi hasil terdiri dari empat komponen: Akurasi menunjukkan prediksi yang benar dari total prediksi, Presisi (P) adalah ukuran kemampuan sistem dalam membedakan antara serangan dan yang dianggap normal, *Recall* atau tingkat positif sebenarnya menunjukkan jumlah serangan DDoS yang diprediksi dalam serangan DDoS aktual, dan Skor F1 didefinisikan sebagai rata-rata harmonis dari recall dan presisi, dengan hasil skor F1 berkisar antara 0 terburuk dan 1 terbaik. Hasil dari studi ini dikelompokkan berdasarkan normal dan tidak normal. Dalam setiap hasil, terdapat empat ekspektasi, yaitu: *True Positive* (TP) adalah pengenalan

serangan DDoS yang akurat; *True Negative* (TN) adalah pengenalan yang tepat terhadap catatan normal; *False Positive* (FP) mengidentifikasi serangan DDoS secara keliru; dan *False Negative* (FN) mengenali catatan normal dengan keliru [52].

3.2.5 Hasil

Pada tahap ini dilakukan untuk menarik kesimpulan yang dari dasil pegolahan dataset dan clasifikasi yang telah di lakukan sebelumnya. Data yang diperoleh tersebut di Setelah melakukan evaluasi akan didapat informasi yang selanjudnya akan dilakukan analisis data untuk menentukan peforma dari algoritma yang di gunakan berdasaarkan seleksi fitur yang di pakai. Dengan melihat tingkat akurasi dan kesalahan dalam pembacaan data yang telah melalui pelatihan dan penguian sebelumnya.

3.3 Timeline penelitian

Dalam studi time line sangat krusial dalam penelitian, karena menggambarkan urutan aktivitas yang akan dilaksanakan. Garis waktu itu sendiri adalah elemen yang menjaga sejarah nilai-nilai dari sub-kumpulan masalah perencanaan tertentu yang jelas terdefinisi. Garis waktu menghitung pernyataan temporal yang dihasilkan, yang memberikan deskripsi singkat mengenai interval waktu di mana status Garis Waktu dipahami, serta sejauh mana status tersebut dipahami [53]. Dalam penelitian deteksi DDOS Attack dilakukan menggunakan timeline sebagai berikut.

No	Kegiatan	2025																									
		Februari				Maret				April				Mei				Juni				Juli				Agustus			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Pra-Penelitian																													
1	survei Literatur																												
2	Identifikasi Masalah																												
3	Studi Pustaka																												
4	Penentuan Sumber Data																												
Bimbingan Proposal																													
Seminar Proposal																													
Kegiatan Penelitian																													
	Pengolahan Data																												
	Analisa Visualisasi Data																												
Penarikan kesimpulan																													
	Penyusunan Tesis																												
Seminar Tesis																													

Gambar 3. 6 Time Line Penelitian