

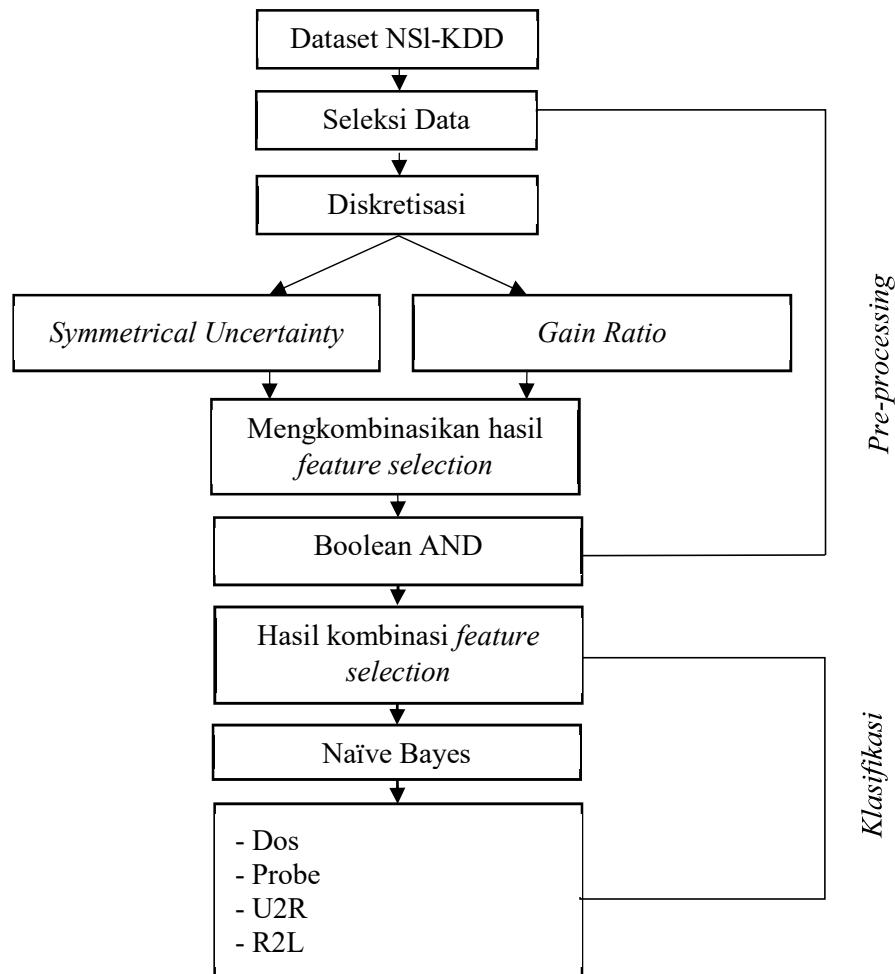
BAB IV

ANALISA DAN PERANCANGAN

Pada bab ini akan membahas tentang analisa data dan perancangan model sistem klasifikasi yang menggunakan metode Naive Bayes Classifier.

4.1 Analisa Data

Pada tahap analisa data, akan dilakukan proses sesuai tahapan *Knowledge Discovery in Database* (KDD). Yaitu proses *data selection*, *preprocessing*, klasifikasi menggunakan metode *Naïve Bayes*.



Gambar 4.1 Flowchart Analisa Data

4.1.1 Data Selection

Tahap *Data selection* pada penelitian ini merupakan tahap memilih data yang akan penulis gunakan dari beberapa pilihan data yang berhasil dikumpulkan. Data yang digunakan adalah dataset NSL-KDD lengkap yang kemudian diambil secara acak dan merata pada setiap kelas sesuai jumlah data terkecil pada kelasnya. Rincian dataset dapat dilihat pada Tabel 4.1.

Tabel 4.1 Dataset NSL-KDD

Type Serangan/ Class	Dataset NSL-KDD
Dos	52
Probe	52
U2R	52
R2L	52
Total	208

4.1.2 Pre-Processing

Pre-processing pada penelitian ini terbagi menjadi tiga tahapan, yang pertama adalah tahap diskretisasi, yang kedua tahap *feature selection* menggunakan algoritma *Gain Ratio* dan *Symmetrical Uncertainty*, dan yang ketiga adalah mengkombinasikan 20 atribut yang memiliki nilai tertinggi hasil dari masing-masing *feature selection* menggunakan pendekatan *Boolean AND Operator*, sehingga terpilihlah atribut yang saling terkait dari setiap algoritma *feature selection*.

4.1.2.1 Discretization

Information gain dan *symmetrical uncertainty* merupakan teknik seleksi fitur yang memakai metode *scoring* untuk data nominal atau atribut kontinu yang ditransformasikan ke dalam bentuk atribut kategorik. Pada penelitian ini akan dilakukan diskretisasi data pada atribut yang memiliki data kontinu ke dalam bentuk kategorik 10 interval menggunakan *tools WEKA 3.8.2*

1. Diskretisasi Atribut *Duration*

Berikut ini, akan ditampilkan cara diskretisasi data pada atribut *duration* menggunakan persamaan 2.11

Diketahui :

$$\text{Nilai minimum (A)} = 0$$

$$\text{Nilai Maximum (B)} = 41285$$

$$\text{Interval (N)} = 10$$

$$\begin{aligned}\text{Jarak Interval (W)} &= (41285 - 0)/10 \\ &= 4128.5\end{aligned}$$

Hasil diskretisasi data pada atribut *duration* dapat dilihat pada tabel 4.2.

Tabel 4.2 Diskretisasi Data Atribut Duration

No	Kategori	Jumlah
1	Inf - 4128.5	197
2	4128.5-8257	4
3	8257-12385.5	2
4	12385.5-16514	1
5	16514-20642.5	0
6	20642.5-24771	0
7	24771-28899.5	1
8	28899.5-33028	1
9	33028-37156.5	1
10	37156.5 - inf	1
Total		208

Hasil diskritisasi untuk atribut *duration* menggunakan tools WEKA 3.8.2 dapat dilihat pada gambar 4.2.

Selected attribute			
Name: duration		Type: Nominal	
Missing: 0 (0%)		Distinct: 8	
No.	Label	Count	Weight
1	'[-inf-4128.5]'	197	197.0
2	'(4128.5-8257]'	4	4.0
3	'(8257-12385.5]'	2	2.0
4	'(12385.5-16514]'	1	1.0
5	'(16514-20642.5]'	0	0.0
6	'(20642.5-24771]'	0	0.0
7	'(24771-28899.5]'	1	1.0
8	'(28899.5-33028]'	1	1.0
9	'(33028-37156.5]'	1	1.0
10	'(37156.5-inf]	1	1.0

Gambar 4.2 Diskritisasi Atribut *Duration* dengan WEKA

2. Diskretisasi Atribut *src_bytes*

Berikut ini, akan ditampilkan diskretisasi data pada atribut *src_bytes* menggunakan persamaan 2.11

Diketahui :

$$\text{Nilai minimum (A)} = 0$$

$$\text{Nilai Maximum (B)} = 5133876$$

$$\text{Interval (N)} = 10$$

$$\begin{aligned} \text{Jarak Interval (W)} &= (5133876 - 0) / 10 \\ &= 513387.6 \end{aligned}$$

Hasil diskretisasi data pada atribut *src_bytes* dapat dilihat pada tabel 4.3

Tabel 4.3 Diskretisasi Data Atribut *src_bytes*

No	Kategori	Jumlah
1	inf-513387.6	205
2	513387.6-1026775.2	0
3	1026775.2-1540162.8	0
4	1540162.8-2053550.4	0
5	2053550.4-2566938	0
6	2566938-3080325.6	0
7	3080325.6-3593713.2	0
8	3593713.2-4107100.8	0
9	4107100.8-4620488.4	0
10	4620488.4-inf	3

Hasil diskritisasi untuk atribut *src_bytes* menggunakan tools WEKA 3.8.2 dapat dilihat pada gambar 4.3.

Selected attribute			
Name:	src_bytes	Type:	Nominal
Missing:	0 (0%)	Distinct:	2
No.	Label	Count	Weight
1	'(-inf-513387.6]'	205	205.0
2	'(513387.6-102677...	0	0.0
3	'(1026775.2-15401...	0	0.0
4	'(1540162.8-20535...	0	0.0
5	'(2053550.4-25669...	0	0.0
6	'(2566938-3080325...	0	0.0
7	'(3080325.6-35937...	0	0.0
8	'(3593713.2-41071...	0	0.0
9	'(4107100.8-46204...	0	0.0
10	'(4620400-46800...	2	2.0

Gambar 4.3 Diskritisasi Atribut Src_bytes dengan WEKA

Hasil Diskretisasi pada atribut lainnya yang memiliki type data kontinu dapat dilihat pada lampiran A.

4.1.2.2 *Feature selection*

Pada bagian ini, akan dibahas tentang perhitungan algoritma seleksi fitur. Seleksi fitur yang digunakan pada penelitian ini adalah *Gain Ratio* dan *Symmetrical Uncertainty*.

1. *Gain Ratio*

Berdasarkan penjelasan yang terdapat pada landasan teori, terdapat empat tahapan sebelum dapat menghitung *Gain Ratio*. Yaitu menghitung *Entropy Data*, *Entropy Atribut*, *Information Gain* dan *Split Info*, dan akan dibahas dibawah ini.

A. Menghitung Entropy Data

Sebelum menghitung entropy data, terlebih dahulu harus diketahui himpunan data pada setiap kelasnya. Himpunan data kelas dapat dilihat pada tabel 4.4.

Tabel 4.4 Himpunan Data Kelas pada Dataset NSL-KDD

Total Himpunan Kasus	Dos	Probe	U2R	R2L
208	52	52	52	52

Untuk menghitung entropy data dapat dilakukan menggunakan persamaan 2.3, perhitungan entropy dapat dilihat sebagai berikut:

$$\begin{aligned}
 \text{Entropy Data} &= (-52 / 208) * \log_2 (52/208) + (-52 / 208) * \log_2 (52/208) + \\
 &\quad (-52 / 208) * \log_2 (52/208) + (-52 / 208) * \log_2 (52/208) \\
 &= 0,5 + 0,5 + 0,5 + 0,5 \\
 &= 2,0
 \end{aligned}$$

B. Menghitung Entropy Atribut

Berikut ini akan ditampilkan perhitungan entropy setiap atribut yang terdapat pada dataset NSL-KDD.

a. Entropy Atribut (*Duration*)

Berikut ini merupakan himpunan data kelas pada atribut *duration* yang dapat dilihat pada tabel 4.5.

Tabel 4.5 Himpunan Data Kelas Atribut Duration

Partisi (Kemunculan Data)	Dos	Probe	R2I	U2R	Himpunan kasus Partisi
inf-4128.5	52	46	52	47	197
12385.5-16514	0	0	0	1	1
24771-28899.5	0	1	0	0	1
28899.5-33028	0	1	0	0	1
33028-37156.5	0	1	0	0	1
4128.5-8257	0	0	0	4	4
8257-12385.5	0	2	0	0	2
37156.5-inf	0	1	0	0	1
Total	52	52	52	52	208

Untuk menghitung entropy atribut *duration*, pertama sekali dilakukan menghitung entropy partisi (kemunculan data) dalam atribut tersebut yang dapat dihitung menggunakan persamaan 2.6.

$$\begin{aligned}
 \text{Entropy}(\text{Duration} | \text{inf-4128.5}) &= (-52/197) * \log_2(52/197) + \\
 &\quad (-46/197) * \log_2(46/197) + \\
 &\quad (-52/197) * \log_2(52/197) + \\
 &\quad (-47/197) * \log_2(47/197) \\
 &= 0.507 + 0.490 + 0.507 + 0.493 \\
 &= \mathbf{1.998}
 \end{aligned}$$

$$\begin{aligned}
 \text{Entropy}(\text{Duration} | 12385.5-16514) &= (-0/1) * \log_2(0/1) + (-0/1) * \log_2(0/1) + \\
 &\quad (-0/1) * \log_2(0/1) + (-1/1) * \log_2(1/1) \\
 &= 0 + 0 + 0 + 0 \\
 &= \mathbf{0}
 \end{aligned}$$

Entropy partisi (kemunculan data) didalam atribut *Duration* secara lengkap dapat dilihat pada tabel 4.6

Tabel 4.6 Entropy Partisi Atribut Duration

Partisi	Entropy
inf-4128.5	1.998
12385.5-16514	0
24771-28899.5	0
28899.5-33028	0
33028-37156.5	0
4128.5-8257	0
8257-12385.5	0
37156.5-inf	0

Setelah Entropy setiap partisi dalam atribut *Duration* didapatkan, selanjutnya masuk ke tahap menghitung Entropy Atribut menggunakan persamaan 2.5

$$\begin{aligned}
 \text{Entropy}(\text{Duration}) &= ((197/208) * 1.998) + ((1/208) * 0) + ((1/208) * 0) + \\
 &\quad ((1/208) * 0) + ((1/208) * 0) + ((4/208) * 0) + \\
 &\quad ((2/208) * 0) + ((1/208) * 0) \\
 &= 1.892 + 0 + 0 + 0 + 0 + 0 + 0 \\
 &= \mathbf{1.892}
 \end{aligned}$$

b. Entropy Atribut (*Protocol_type*)

Berikut ini merupakan himpunan data kelas pada atribut *Protocol_type* yang dapat dilihat pada tabel 4.7

Tabel 4.7 Himpunan Data Kelas Atribut *Protocol_type*

Partisi (Kemunculan Data)	Dos	Probe	R2l	U2R	Himpunan kasus Partisi
tcp	49	28	49	52	178
icmp	0	21	0	0	21
udp	3	3	3	0	9
Total	52	52	52	52	208

Untuk menghitung entropy atribut *Protocol_type*, pertama sekali dilakukan menghitung entropy partisi (kemunculan data) dalam atribut tersebut, yang dapat dihitung menggunakan persamaan 2.6.

$$\begin{aligned}
 \text{Entropy}(\text{Protocol_type} | \text{tcp}) &= (-49/178)*\log_2(49/178) + \\
 &\quad (-28/178)*\log_2(28/178) + \\
 &\quad (-49/178)*\log_2(49/178) + \\
 &\quad (-52/178)*\log_2(52/178) \\
 &= 0.499 + 0.420 + 0.512 + 0.519 \\
 &= 1.950
 \end{aligned}$$

$$\begin{aligned}
 \text{Entropy}(\text{Protocol_type} | \text{icmp}) &= (-0/21)*\log_2(0/21) + (-21/21)*\log_2(21/21) + \\
 &\quad (-0/21)*\log_2(0/21) + (-0/21)*\log_2(0/21) \\
 &= 0 + 0 + 0 + 0 \\
 &= 0
 \end{aligned}$$

$$\begin{aligned}
 \text{Entropy}(\text{Protocol_type} | \text{udp}) &= (-3/9)*\log_2(3/9) + (-3/9)*\log_2(3/9) + \\
 &\quad (-3/9)*\log_2(3/9) + (-0/9)*\log_2(0/9) \\
 &= 0.092 + 0.528 + 0.528 + 0 \\
 &= 1.149
 \end{aligned}$$

Entropy partisi pada atribut *Protocol_type* secara lengkap dapat dilihat pada tabel 4.8.

Tabel 4.8 Entropy Partisi Atribut *Protocol_type*

Partisi (Kemunculan Data)	Entropy
tcp	1.950
icmp	0
udp	1.149

Setelah Entropy setiap partisi dalam atribut *Protocol_type* didapatkan, selanjutnya masuk ke tahap menghitung Entropy Atribut menggunakan persamaan 2.5

$$\begin{aligned}
 \text{Entropy}(\text{Protocol_type}) &= ((178/208) * 1.950) + ((21/208) * 0) + \\
 &\quad ((9/208) * 1.149) \\
 &= 1.67986 + 0 + 0.06858 \\
 &= \mathbf{1.748}
 \end{aligned}$$

Untuk menghitung Entropy Atribut lainnya dilakukan dengan cara yang sama seperti menghitung entropy atribut *duration* dan *protocol_type*. Hasil perhitungan entropy atribut lainnya secara lengkap dapat dilihat pada tabel 4.9

Tabel 4.9 Hasil Perhitungan Entropy Semua Atribut

No	Atribut	Entropy
1	duration	1.892
2	protocol_type	1.748
3	service	0.689
4	flag	1.151
5	src_bytes	1.971
6	dst_bytes	1.951
7	Land	2
8	wrong_fragment	1.971
9	urgent	1.990
10	hot	1.639
11	num_failed_logins	1.984
12	logged_in	1.214
13	num_compromised	1.900

No	Atribut	Entropy
14	root_shell	1.727
15	su_attempted	2
16	num_root	1.971
17	num_file_creations	1.773
18	num_shells	1.957
19	num_access_files	1.981
20	num_outbound_cmds	2
21	is_host_login	2
22	is_guest_login	1.796
23	count	1.264
24	srv_count	1.703
25	serror_rate	1.633
26	srv_serror_rate	1.595
27	rror_rate	1.777
28	srv_rror_rate	1.818
29	same_srv_rate	1.413
30	diff_srv_rate	1.881
31	srv_diff_host_rate	1.827
32	dst_host_count	1.537
33	dst_host_srv_count	1.746
34	dst_host_same_srv_rate	1.455
35	dst_host_diff_srv_rate	1.700
36	dst_host_same_src_port_rate	1.465
37	dst_host_srv_diff_host_rate	1.553
38	dst_host_serror_rate	1.579
39	dst_host_srv_serror_rate	1.596
40	dst_host_rror_rate	1.692
41	dst_host_srv_rror_rate	1.729

C. Menghitung *Information Gain*

Pada bagian ini, akan dibahas tentang menghitung *Information Gain* setiap atribut yang terdapat pada dataset NSL-KDD, *Information Gain* dapat dihitung menggunakan persamaan 2.2

$$\begin{aligned} \text{Information Gain (Duration)} &= 2 - 1.892 \\ &= \mathbf{0.108} \end{aligned}$$

$$\begin{aligned} \text{Information Gain (protocol_type)} &= 2 - 1.748 \\ &= \mathbf{0.252} \end{aligned}$$

$$\begin{aligned} \text{Information Gain (service)} &= 2 - 0.689 \\ &= \mathbf{1.311} \end{aligned}$$

Hasil perhitungan *Information Gain* atribut pada dataset NSL-KDD secara lengkap dapat dilihat pada tabel 4.10

Tabel 4.10 Hasil Perhitungan *Information Gain* pada Atribut

No	Atribut	Information Gain
1	duration	0.108
2	protocol_type	0.252
3	service	1.311
4	flag	0.849
5	src_bytes	0.029
6	dst_bytes	0.049
7	Land	0.000
8	wrong_fragment	0.029
9	urgent	0.010
10	hot	0.361
11	num_failed_logins	0.016
12	logged_in	0.786
13	num_compromised	0.100
14	root_shell	0.273
15	su_attempted	0.000
16	num_root	0.029
17	num_file_creations	0.227
18	num_shells	0.043
19	num_access_files	0.019
20	num_outbound_cmds	0.000
21	is_host_login	0.000
22	is_guest_login	0.204
23	count	0.736
24	srv_count	0.297
25	serror_rate	0.367
26	srv_serror_rate	0.405
27	error_rate	0.223
28	srv_error_rate	0.182
29	same_srv_rate	0.587
30	diff_srv_rate	0.119
31	srv_diff_host_rate	0.173
32	dst_host_count	0.463
33	dst_host_srv_count	0.254
34	dst_host_same_srv_rate	0.545
35	dst_host_diff_srv_rate	0.300
36	dst_host_same_src_port_rate	0.535
37	dst_host_srv_diff_host_rate	0.447
38	dst_host_serror_rate	0.421

No	Atribut	<i>Information Gain</i>
39	dst_host_srv_serror_rate	0.404
40	dst_host_rerror_rate	0.308
41	dst_host_srv_rerror_rate	0.271

D. Menghitung *Split Info*

Pada bagian ini akan ditampilkan cara menghitung *Split Info* dari masing-masing atribut yang terdapat pada Dataset NSL-KDD, *Split Info* dapat dihitung menggunakan persamaan 2.6.

$$\begin{aligned}
 \text{Split Info (Duration)} &= (-197/208) * \log_2(197/208) + (-1/208) * \log_2(1/208) + \\
 &\quad (-1/208) * \log_2(1/208) + (-1/208) * \log_2(1/208) + \\
 &\quad (-1/208) * \log_2(1/208) + (-4/208) * \log_2(4/208) + \\
 &\quad (-2/208) * \log_2(2/208) + (-1/208) * \log_2(1/208) \\
 &= 0.074 + 0.037 + 0.037 + 0.037 + 0.037 + 0.110 + \\
 &\quad 0.064 + 0.037 \\
 &= \mathbf{0.433}
 \end{aligned}$$

$$\begin{aligned}
 \text{Split Info (protocol_type)} &= (-178/208) * \log_2(178/208) + (-21/208) * \log_2(21/208) + \\
 &\quad (-9/208) * \log_2(9/208) \\
 &= 0.192 + 0.334 + 0.196 \\
 &= \mathbf{0.722}
 \end{aligned}$$

Hasil perhitungan *Split Info* pada setiap atribut secara lengkap dapat dilihat pada tabel 4.11

Tabel 4.11 Hasil Perhitungan *Split Info* Pada Atribut

No	Atribut	<i>Split Info</i>
1	duration	0.433
2	protocol_type	0.722
3	service	3.292
4	flag	1.649
5	src_bytes	0.109
6	dst_bytes	0.244
7	Land	0.000
8	wrong_fragment	0.109
9	urgent	0.044

No	Atribut	Split Info
10	hot	0.936
11	num failed logins	0.109
12	logged_in	0.995
13	num compromised	0.301
14	root_shell	0.557
15	su attempted	0.000
16	num root	0.132
17	num file creations	0.683
18	num shells	0.215
19	num access_files	0.122
20	num outbound_cmds	0.000
21	is_host_login	0.000
22	is_guest_login	0.441
23	count	1.595
24	srv_count	0.950
25	serror_rate	0.781
26	srv_serror_rate	0.761
27	rerror_rate	0.882
28	srv_rerror_rate	0.735
29	same_srv_rate	1.475
30	diff_srv_rate	0.545
31	srv_diff_host_rate	0.527
32	dst_host_count	1.377
33	dst_host_srv_count	1.525
34	dst_host_same_srv_rate	1.736
35	dst_host_diff_srv_rate	1.234
36	dst_host_same_src_port_rate	1.861
37	dst_host_srv_diff_host_rate	1.398
38	dst_host_serror_rate	0.869
39	dst_host_srv_serror_rate	0.793
40	dst_host_rerror_rate	1.438
41	dst_host_srv_rerror_rate	0.932

E. Menghitung Gain Ratio

Setelah di dapatkan nilai *Information Gain* dan *Split Info* dari masing-masing atribut yang telah dihitung pada tahap sebelumnya, selanjutnya dilakukan perhitungan *Gain Ratio* menggunakan persamaan 2.1

$$\begin{aligned}
 \text{Gain Ratio (duration)} &= 0.108 / 0.433 \\
 &= \mathbf{0.249}
 \end{aligned}$$

$$Gain\ Ratio (protocol_type) = 0.252 / 0.722$$

$$= \mathbf{0.348}$$

$$Gain\ Ratio (service) = 1.311 / 3.291$$

$$= \mathbf{0.398}$$

Hasil perhitungan *Gain ratio* pada setiap atribut secara lengkap dapat dilihat pada tabel 4.12

Tabel 4.12 Hasil Perhitungan *Gain Ratio* Pada Atribut

No	Atribut	Gain Ratio
1	duration	0.249
2	protocol_type	0.348
3	service	0.398
4	flag	0.515
5	src_bytes	0.269
6	dst_bytes	0.202
7	Land	0
8	wrong_fragment	0.269
9	urgent	0.220
10	hot	0.385
11	num_failed_logins	0.145
12	logged_in	0.790
13	num_compromised	0.333
14	root_shell	0.490
15	su_attempted	0
16	num_root	0.220
17	num_file_creations	0.333
18	num_shells	0.201
19	num_access_files	0.159
20	num_outbound_cmds	0
21	is_host_login	0
22	is_guest_login	0.463
23	count	0.461
24	srv_count	0.313
25	serror_rate	0.470
26	srv_serror_rate	0.533
27	rerror_rate	0.253
28	srv_rerror_rate	0.248

No	Atribut	Gain Ratio
29	same_srv_rate	0.398
30	diff_srv_rate	0.219
31	srv_diff_host_rate	0.328
32	dst_host_count	0.336
33	dst_host_srv_count	0.167
34	dst_host_same_srv_rate	0.314
35	dst_host_diff_srv_rate	0.243
36	dst_host_same_src_port_rate	0.288
37	dst_host_srv_diff_host_rate	0.320
38	dst_host_serror_rate	0.485
39	dst_host_srv_serror_rate	0.509
40	dst_host_rerror_rate	0.214
41	dst_host_srv_rerror_rate	0.291

Setelah nilai *Gain Ratio* dari setiap atribut didapatkan, selanjutnya dilakukan pengurutan berdasarkan nilai terbesar ke nilai terkecil. 20 atribut dengan nilai terbaik dapat dilihat pada tabel 4.13.

Tabel 4.13 Atribut Terbaik Hasil Perhitungan *Gain Ratio*

No Atribut	Atribut	Gain Ratio
12	logged_in	0.79
26	srv_serror_rate	0.533
4	flag	0.515
39	dst_host_srv_serror_rate	0.509
14	root_shell	0.49
38	dst_host_serror_rate	0.485
25	serror_rate	0.47
22	is_guest_login	0.463
23	count	0.461
3	service	0.398
29	same_srv_rate	0.398
10	hot	0.385
2	protocol_type	0.348
32	dst_host_count	0.336
13	num_compromised	0.333
17	num_file_creations	0.333
31	srv_diff_host_rate	0.328

No Atribut	Atribut	Gain Ratio
37	dst_host_srv_diff_host_rate	0.32
34	dst_host_same_srv_rate	0.314
24	srv_count	0.313

2. *Symmetrical Uncertainty*

Dalam menghitung nilai *Symmetrical Uncertainty*, yang pertama harus dilakukan adalah menghitung *Entropy Data*, *Information Gain* dan *Split Info* seperti yang telah dibahas pada tahapan menghitung *Gain Ratio*. Untuk menghitung *Symmetrical Uncertainty* maka dapat dilakukan menggunakan persamaan 2.9.

$$\begin{aligned}
 \text{Symmetrical Uncertainty (duration)} &= 2 * (0.108 / (2 + 0.433)) \\
 &= 2 * (0.108 / 2.433) \\
 &= 2 * 0.044 \\
 &= \mathbf{0.089} \\
 \text{Symmetrical Uncertainty (protocol_type)} &= 2 * (0.252 / (2 + 0.722)) \\
 &= 2 * (0.252 / 2.722) \\
 &= 2 * 0.092 \\
 &= \mathbf{0.185} \\
 \text{Symmetrical Uncertainty (service)} &= 2 * (1.311 / (2 + 3.292)) \\
 &= 2 * (1.311 / 5.292) \\
 &= 2 * 0.247 \\
 &= \mathbf{0.496}
 \end{aligned}$$

Hasil perhitungan *Symmetrical Uncertainty* pada semua atribut secara lengkap dapat dilihat pada tabel 4.14.

Tabel 4.14 Hasil Perhitungan *Symmetrical Uncertainty*

No	Atribut	Symmetrical Uncertainty
1	duration	0.089
2	protocol_type	0.185
3	service	0.496
4	flag	0.465
5	src_bytes	0.028

No	Atribut	<i>Symmetrical Uncertainty</i>
6	dst_bytes	0.044
7	Land	0.000
8	wrong_fragment	0.028
9	urgent	0.009
10	hot	0.246
11	num_failed_logins	0.015
12	logged_in	0.525
13	num_compromised	0.087
14	root_shell	0.213
15	su_attempted	0.000
16	num_root	0.027
17	num_file_creations	0.169
18	num_shells	0.039
19	num_access_files	0.018
20	num_outbound_cmds	0.000
21	is_host_login	0.000
22	is_guest_login	0.167
23	count	0.409
24	srv_count	0.201
25	serror_rate	0.264
26	srv_serror_rate	0.294
27	rerror_rate	0.155
28	srv_rerror_rate	0.133
29	same_srv_rate	0.338
30	diff_srv_rate	0.094
31	srv_diff_host_rate	0.137
32	dst_host_count	0.274
33	dst_host_srv_count	0.144
34	dst_host_same_srv_rate	0.292
35	dst_host_diff_srv_rate	0.185
36	dst_host_same_src_port_rate	0.277
37	dst_host_srv_diff_host_rate	0.263
38	dst_host_serror_rate	0.294
39	dst_host_srv_serror_rate	0.289
40	dst_host_rerror_rate	0.179
41	dst_host_srv_rerror_rate	0.185

Setelah nilai *Symmetrical Uncertainty* dari setiap atribut didapatkan, selanjutnya dilakukan pengurutan berdasarkan nilai terbesar ke nilai terkecil. 20 atribut dengan nilai terbaik dapat dilihat pada tabel 4.15.

Tabel 4.15 Atribut Terbaik Hasil Perhitungan *Symmetrical Uncertainty*

No Atribut	Atribut	<i>Symmetrical Uncertainty</i>
12	logged_in	0.525
3	service	0.496
4	flag	0.465
23	count	0.409
29	same_srv_rate	0.338
26	srv_serror_rate	0.294
38	dst_host_serror_rate	0.294
34	dst_host_same_srv_rate	0.292
39	dst_host_srv_serror_rate	0.289
36	dst_host_same_src_port_rate	0.277
32	dst_host_count	0.274
25	serror_rate	0.264
37	dst_host_srv_diff_host_rate	0.263
10	hot	0.246
14	root_shell	0.213
24	srv_count	0.201
2	protocol_type	0.185
35	dst_host_diff_srv_rate	0.185
41	dst_host_srv_error_rate	0.185
40	dst_host_error_rate	0.179

4.1.2.3 Boolean AND Operator

Pada bagian ini, 20 atribut dengan nilai tertinggi hasil dari *Gain Ration* dan *Symmetrical Uncertainty* dipilih, kemudian dilakukan kombinasi menggunakan *Boolean AND Operator*. 20 atribut dengan nilai tertinggi hasil dari *Gain Ratio* dan *Symmetrical Uncertainty* dapat dilihat pada tabel 4.16

Tabel 4.16 Atribut Terbaik Hasil Perhitungan GR dan SU

Feature Selection	20 Atribut Hasil Feature Selection
GR	12,26,4,39,14,38,25,22,23,29,3,10,2,32,13,17,31,37,34,24
SU	12,3,4,23,29,38,26,34,39,36,32,25,37,10,14,24,35,41,2,40

Selanjutnya hasil *feature selection* GR dan SU pada tabel 4.16 dikombinasikan menggunakan persamaan 2.7, dan didapatkan 16 atribut yang terpilih yang dapat dilihat pada Tabel 4.17.

Tabel 4.17 Hasil Boolean AND 20 Atribut GR dan SU

Feature Selection	Hasil Kombinasi Feature Selection GR dan SU
GR&SU	2,3,4,10,12,14,23,24,25,26,29,32,34,37,38,39

4.1.3 Analisa Model Klasifikasi Naïve Bayes

Setelah melewati proses pemilihan atribut, pada bagian ini akan dilakukan analisa data menggunakan model klasifikasi Naïve Bayes. Analisa model klasifikasi Naïve Bayes akan dilakukan sebanyak 3 kali, yaitu dengan jumlah data latih dan uji sebanyak 70:10, 80:10 dan 90:10 yang dibagi secara acak dari dataset NSL-KDD.

Berikut ini akan ditampilkan analisa model klasifikasi Naïve Bayes pada data latih dan Uji 70:10, terdapat 4 proses utama dalam mengklasifikasi data. yaitu : mencari probabilitas kelas, mencari probabilitas atribut pada setiap kelas, mencari probabilitas akhir, dan prediksi.

Tabel 4.18 Dataset Latih dan Uji (70:10)

Class	Data Latih 70%	Data Uji 10%
dos	37	5
r2l	37	5
probe	36	5
u2r	36	5
Total	146	20

4.1.4.1 Mencari Probabilitas Kelas

Pada bagian ini akan ditampilkan perhitungan untuk mencari probabilitas kelas, Probabilitas kelas dapat dihitung menggunakan persamaan 2.15

$$P(\text{dos}) = \frac{37}{146} = 0.25$$

$$P(\text{r2l}) = \frac{37}{146} = 0.25$$

$$P(\text{probe}) = \frac{36}{146} = 0.25$$

$$P(\text{u2r}) = \frac{36}{146} = 0.25$$

Probabilitas kelas selengkapnya dapat dilihat pada Tabel 4.19:

Tabel 4.19 Probabilitas Kelas

No	Kelas	Probabilitas P(Y)
1	dos	0.25
2	r2l	0.25
3	probe	0.25
4	u2r	0.25

4.1.4.2 Mencari Probabilitas Atribut pada setiap kelas

Pada tahap ini akan ditampilkan cara menghitung probabilitas atribut yang dilakukan pada dataset latih. Berhubungan dengan adanya nilai 0 (nol) untuk jumlah sampel dalam kelas dari suatu atribut, akan menyebabkan nilai probabilitas akan menjadi nol, maka untuk mengatasi hal tersebut bisa dilakukan menggunakan persamaan *Laplace Smoothing*.

1. Probabilitas Atribut *Protocol_type*

Sebelum menghitung probabilitas atribut, harus diketahui jumlah data setiap partisi (kemunculan data) pada setiap kelasnya. Data partisi pada atribut *protocol_type* dapat dilihat pada tabel 4.20.

Tabel 4.20 Partisi Atribut *protocol_type*

Partisi (Kemunculan data)	DoS	R2L	Probe	U2R
tcp	35	37	21	33
icmp	0	0	14	0
udp	2	0	1	3
Total	37	37	36	36

Berikut ini akan ditampilkan perhitungan probabilitas atribut *protocol_type* yang dihitung menggunakan persamaan 2.18.

$$P(\text{protocol_type} = \text{tcp} | \text{class} = \text{"dos"}) = 35+1 / 37+1 = 35.027$$

$$P(\text{protocol_type} = \text{tcp} | \text{class} = \text{"r2l"}) = 37+1 / 37+1 = 38.027$$

$$P(\text{protocol_type} = \text{tcp} | \text{class} = \text{"probe"}) = 21+1 / 36+1 = 22.028$$

$$P(\text{protocol_type} = \text{tcp} | \text{class} = \text{"u2r"}) = 33+1 / 36+1 = 34.028$$

$$P(\text{protocol_type} = \text{icmp} | \text{class} = \text{"dos"}) = 0+1 / 37+1 = 1.027$$

$$P(\text{protocol_type} = \text{icmp} | \text{class} = \text{"r2l"}) = 0+1 / 37+1 = 1.027$$

$$P(\text{protocol_type} = \text{icmp} | \text{class} = \text{"probe"}) = 14+1 / 36+1 = 15.028$$

$$P(\text{protocol_type} = \text{icmp} | \text{class} = \text{"u2r"}) = 0+1 / 36+1 = 1.028$$

$$P(\text{protocol_type} = \text{udp} | \text{class} = \text{"dos"}) = 2+1 / 37+1 = 4.027$$

$$P(\text{protocol_type} = \text{udp} | \text{class} = \text{"r2l"}) = 0+1 / 37+1 = 1.027$$

$$P(\text{protocol_type} = \text{udp} | \text{class} = \text{"probe"}) = 1+1 / 36+1 = 2.028$$

$$P(\text{protocol_type} = \text{udp} | \text{class} = \text{"u2r"}) = 3+1 / 36+1 = 4.028$$

Probabilitas Atribut *protocol_type* secara lengkap dapat dilihat pada tabel 4.21

Tabel 4.21 Probabilitas Atribut *Protocol_type*

Partisi (Kemunculan data)	DoS	R2L	Probe	U2R
tcp	35.027	38.027	22.028	34.028
icmp	1.027	1.027	15.028	1.028
udp	4.027	1.027	2.028	4.028

2. Probabilitas Atribut Service

Sebelum menghitung probabilitas atribut, harus diketahui jumlah data setiap partisi (kemunculan data) pada setiap kelasnya. Data partisi pada atribut *service* dapat dilihat pada tabel 4.22

Tabel 4.22 Partisi Atribut Service

Partisi (Kemunculan data)	DoS	R2L	Probe	U2R
private	13	0	13	0
ftp_data	2	19	0	8
telnet	0	3	0	24
eco_i	0	0	14	0
ftp	0	14	0	1
remote_job	1	0	0	0
netbios_dgm	1	0	1	0
name	1	0	0	0
gopher	0	0	1	0
netbios_ns	1	0	0	0
other	0	0	5	3
mtp	2	0	0	0
finger	2	0	0	0
supdup	2	0	0	0
uucp_path	2	0	0	0
Z39_50	1	0	0	0
csnet_ns	2	0	0	0
sql_net	0	0	1	0
uucp	1	0	0	0
time	0	0	1	0
http	1	0	0	0
auth	1	0	0	0
domain	1	0	0	0
bgp	2	0	0	0
imap4	0	1	0	0
Idap	1	0	0	0
Total	37	37	36	36

Sebagai contoh, penulis akan menampilkan perhitungan probabilitas atribut “service” pada dua partisi (kemunculan data) didalam atribut **service** yaitu: “private” dan “ftp_data” yang akan dihitung menggunakan persamaan 2.18.

$$P(\text{services} = \text{private} | \text{class} = \text{"dos"}) = 13+1 / 37+1 = 14.027$$

$$P(\text{services} = \text{private} | \text{class} = \text{"r2l"}) = 0+1 / 37+1 = 1.027$$

$$P(\text{services} = \text{private} | \text{class} = \text{"probe"}) = 13+1 / 36+1 = 14.028$$

$$P(\text{services} = \text{private} | \text{class} = \text{"u2r"}) = 0+1 / 36+1 = 1.028$$

$$P(\text{services} = \text{ftp_data} | \text{class} = \text{"dos"}) = 2+1 / 37+1 = 3.027$$

$$P(\text{services} = \text{ftp_data} | \text{class} = \text{"r2l"}) = 19+1 / 37+1 = 20.027$$

$$P(\text{services} = \text{ftp_data} | \text{class} = \text{"probe"}) = 0+1 / 36+1 = 1.028$$

$$P(\text{services} = \text{ftp_data} | \text{class} = \text{"u2r"}) = 8+1 / 36+1 = 9.028$$

Hasil dari probabilitas atribut *service* pada setiap partisi (kemunculan data) dan setiap kelasnya dapat dilihat pada Tabel 4.23

Tabel 4.23 Probabilitas Atribut service

Partisi (Kemunculan data)	DoS	R2L	Probe	U2R
private	14.027	1.027	14.028	1.028
ftp_data	3.027	20.027	1.028	9.028
telnet	1.027	4.027	1.028	25.028
eco_i	1.027	1.027	15.028	1.028
ftp	1.027	15.027	1.028	2.028
remote_job	2.027	1.027	1.028	1.028
netbios_dgm	2.027	1.027	2.028	1.028
name	2.027	1.027	1.028	1.028
gopher	1.027	1.027	2.028	1.028
netbios_ns	2.027	1.027	1.028	1.028
other	1.027	1.027	6.028	4.028
mtp	3.027	1.027	1.028	1.028
finger	3.027	1.027	1.028	1.028
supdup	3.027	1.027	1.028	1.028
uucp_path	3.027	1.027	1.028	1.028

Partisi (Kemunculan data)	DoS	R2L	Probe	U2R
Z39_50	2.027	1.027	1.028	1.028
csnet_ns	3.027	1.027	1.028	1.028
sql_net	1.027	1.027	2.028	1.028
uucp	2.027	1.027	1.028	1.028
time	1.027	1.027	2.028	1.028
http	2.027	1.027	1.028	1.028
auth	2.027	1.027	1.028	1.028
domain	2.027	1.027	1.028	1.028
bgp	3.027	1.027	1.028	1.028
imap4	1.027	2.027	1.028	1.028
Idap	2.027	1.027	1.028	1.028

Hasil perhitung probabilitas atribut lainnya dapat dilihat pada lampiran B.

4.1.4.3 Menghitung Probabilitas Akhir

Pada bagian ini akan dilakukan menghitung nilai probabilitas akhir menggunakan dataset uji. Sebelum menghitung probabilitas akhir, langkah pertama yang harus dilakukan adalah melihat partisi (kemunculan data) pada setiap atribut dari data uji, kemudian mengambil nilai probabilitas paritisi tersebut pada setiap kelasnya dari perhitungan probabilitas atribut diatas. Langkah selanjutnya menghitung probabilitas akhir dapat dicari menggunakan persamaan 2.14. Probabilitas akhir dapat dilihat pada tabel 4.24

Tabel 4.24 Probabilitas akhir

Data Uji	DoS	R2L	Probe	U2R
Ke-1	6.59035E+21	829915447.9	5.60719E+13	10226711617
Ke-2	7.08402E+12	1.31063E+21	5.62386E+16	8.10905E+18
Ke-3	3.47881E+11	6.16151E+20	1.39591E+17	2.59274E+20
Ke-4	1.73671E+11	4.97102E+16	6.59446E+20	3.84488E+16
Ke-5	1.46879E+20	1.2557E+15	1.3881E+20	8.26701E+15
Ke-6	2.55898E+13	2.19473E+23	1.93331E+19	6.46172E+22
Ke-7	3.78206E+14	9.1699E+17	1.70264E+16	1.76594E+19
Ke-8	2.64396E+17	4.64726E+18	1.00228E+22	1.46539E+19
Ke-9	1.45646E+21	23019552.57	6.68473E+13	283862212.9

Data Uji	DoS	R2L	Probe	U2R
Ke-10	2.55898E+13	2.19473E+23	1.93331E+19	6.46172E+22
Ke-11	2.55898E+13	2.19473E+23	1.93331E+19	6.46172E+22
Ke-12	2.54572E+11	5.28223E+18	9.54998E+20	2.73296E+18
Ke-13	3.88528E+19	1.2557E+15	1.77715E+19	4.19013E+15
Ke-14	8.68224E+12	3.9085E+21	5.13542E+16	7.30039E+18
Ke-15	2.34488E+11	5.80133E+21	5.36627E+17	9.32476E+22
Ke-16	11869607745	4.97102E+16	2.21168E+20	1.13268E+17
Ke-17	3.05392E+22	829915447.9	7.65306E+14	10226711617
Ke-18	2.55898E+13	2.19473E+23	1.93331E+19	6.46172E+22
Ke-19	6333018876	3.17591E+20	1.48951E+16	1.28155E+23
Ke-20	1.81003E+17	1.67623E+18	9.34365E+20	4.72004E+18

4.1.4.4 Prediksi

Prediksi merupakan kelas dengan nilai probabilitas akhir tertinggi. Hasil prediksi pada data uji dapat dilihat pada Tabel 4.25

Tabel 4.25 Kelas Prediksi

No	Data Uji	Kelas Awal	Kelas Prediksi
1	Ke-1	dos	dos
2	Ke-2	r2l	r2l
3	Ke-3	u2r	r2l
4	Ke-4	Probe	probe
5	Ke-5	dos	dos
6	Ke-6	r2l	r2l
7	Ke-7	u2r	u2r
8	Ke-8	Probe	probe
9	Ke-9	dos	dos
10	Ke-10	r2l	r2l
11	Ke-11	u2r	r2l
12	Ke-12	Probe	probe
13	Ke-13	dos	dos
14	Ke-14	r2l	r2l
15	Ke-15	u2r	u2r
16	Ke-16	Probe	probe
17	Ke-17	dos	dos
18	Ke-18	r2l	r2l

No	Data Uji	Kelas Awal	Kelas Prediksi
19	Ke-19	u2r	u2r
20	Ke-20	Probe	probe

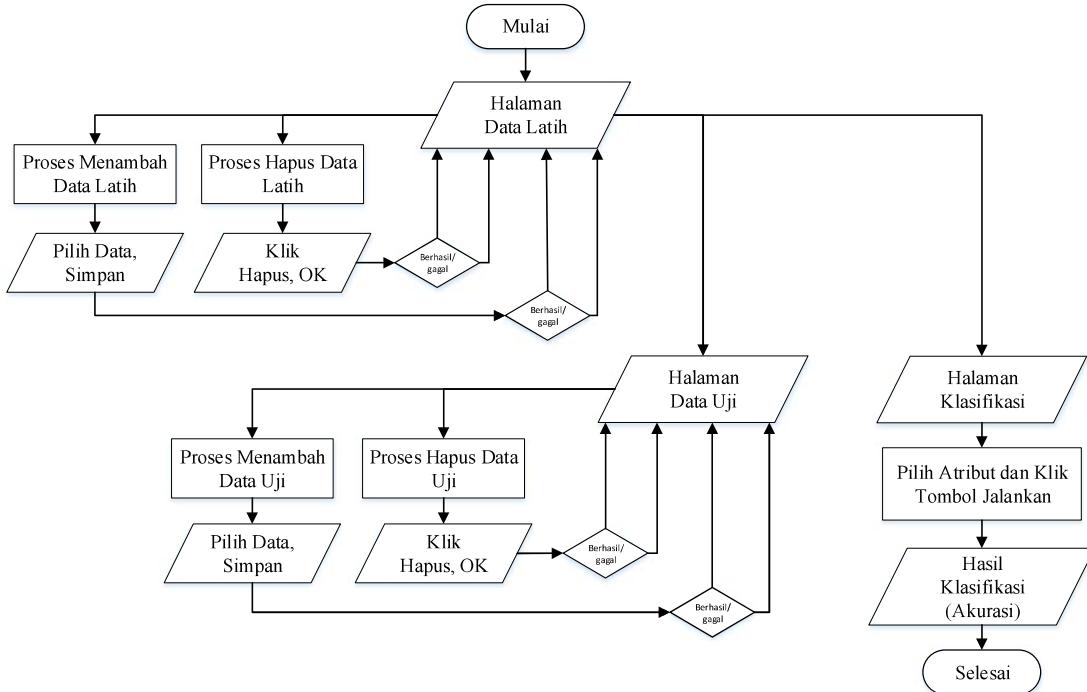
Selanjutnya akan dilakukan analisa terhadap data latih dan data uji yang dibagi menjadi 80:10 dan 90:10 yang dapat dilihat pada lampiran **C** dan **D**.

4.2 Perancangan Sistem

Bagian ini merupakan pembahasan tentang rancangan sistem yang akan dibangun, perancangan pada bagian ini meliputi perancangan fungsional sistem, perancangan database, perancangan struktur menu dan perancangan interface (antarmuka) pada sistem.

4.2.1 Perancangan Fungsional Sistem

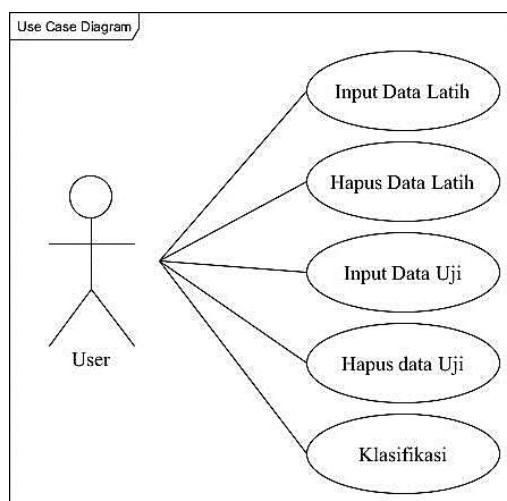
Adapun analisa fungsional sistem dalam membangun sistem klasifikasi dataset NSL-KDD pada penelitian ini akan dilakukan menggunakan UML (*Unified Modeling Language*). UML terdiri dari 4 jenis diagram yaitu *Usecase Diagram*, *Activity Diagram*, *Sequence Diagram*, *Class Diagram*. *Usecase Diagram* merupakan diagram yang menunjukkan hubungan antara actor dan permasalahan yang ada. *Usecase* digunakan untuk analisis dan *desain system*. *Usecase* dapat menggambarkan interaksi antara pengguna sistem itu sendiri, dengan memberikan narasi terhadap sistem yang digunakan. Adapun *Flowchart* system yang akan dibangun dapat dilihat pada gambar 4.4 :



Gambar 4.4 Flowchart System

4.2.1.1 Usecase Diagram

Use case diagram digunakan untuk menggambarkan secara ringkas siapa saja aktor yang berada pada sistem dan apa saja yang bisa dilakukannya, *usecase diagram* untuk sistem yang akan dibuat pada penelitian ini dapat dilihat pada gambar 4.5.



Gambar 4.5 Usecase Diagram System

4.2.1.2 Usecase Spesifikasi

Penjelasan lebih rinci mengenai *usecase diagram* dapat dilihat pada tabel 4.26.

Tabel 4.26 Usecase Spesifikasi

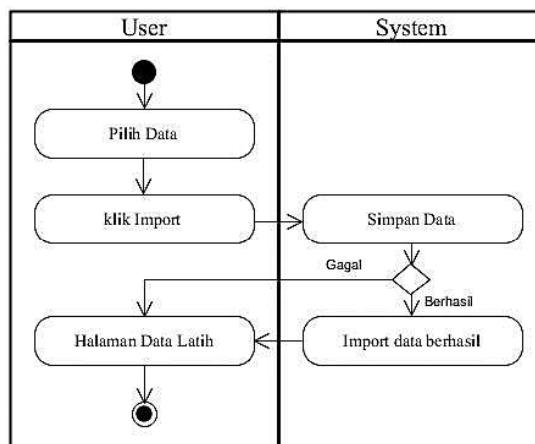
No	Usecase Diagram	Keterangan
1	Input Data Latih	Merupakan menu pengguna untuk melakukan import/input data latih kedalam sistem
2	Hapus Data Latih	Merupakan Menu pengguna untuk melakukan hapus data latih yang telah diinput
3	Input Data Uji	Merupakan menu pengguna untuk melakukan import/input data Uji kedalam sistem
4	Hapus Data Uji	Merupakan Menu pengguna untuk melakukan hapus data Uji yang telah diinput kedalam sistem
5	Klasifikasi	Merupakan Menu pengguna untuk melakukan Proses Klasifikasi pada sistem.

4.2.1.3 Activity Diagram

Activity Diagram merupakan suatu diagram yang digunakan untuk menggambarkan aliran kerja atau aktivitas dari sebuah sistem atau proses bisnis.

1. Activity Diagram Input Data Latih

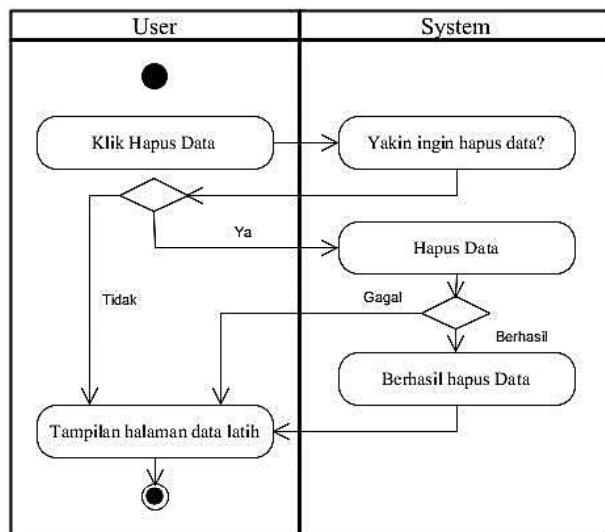
Activity Diagram input data latih untuk sistem yang akan dibuat pada penelitian ini dapat dilihat pada gambar 4.6.



Gambar 4.6 Activity Diagram Input Data Latih

2. Activity Diagram Hapus Data Latih

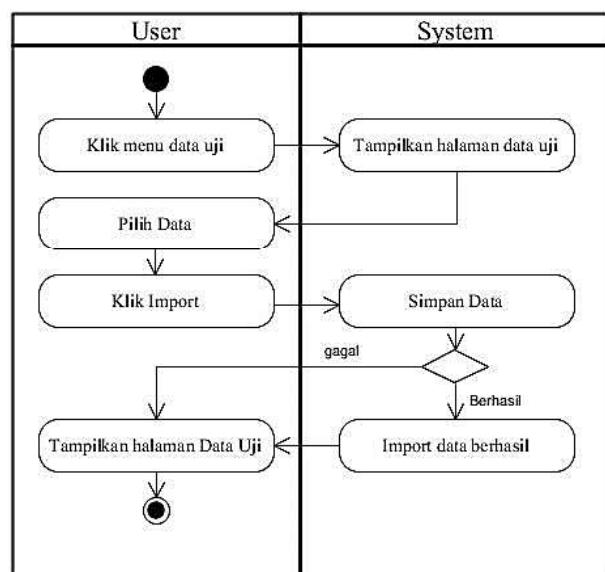
Activity Diagram hapus data latih untuk sistem yang akan dibuat pada penelitian ini dapat dilihat pada gambar 4.7.



Gambar 4.7 Activity Diagram Hapus Data Latih

3. Activity Diagram Input Data Uji

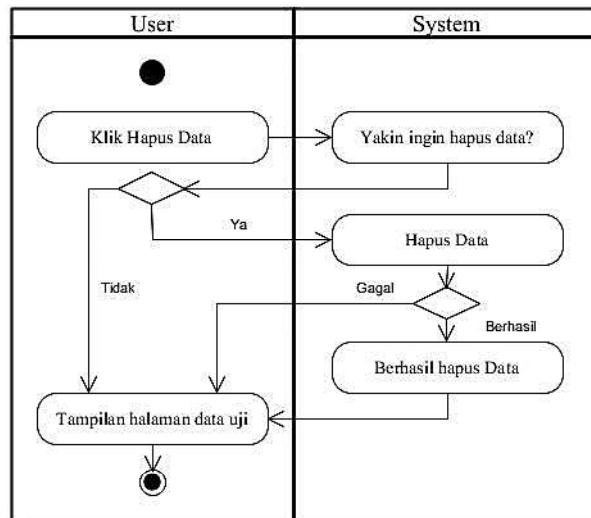
Activity Diagram input data uji untuk sistem yang akan dibuat pada penelitian ini dapat dilihat pada gambar 4.8.



Gambar 4.8 Activity Diagram Input Data Uji

4. Activity Diagram Hapus Data Uji

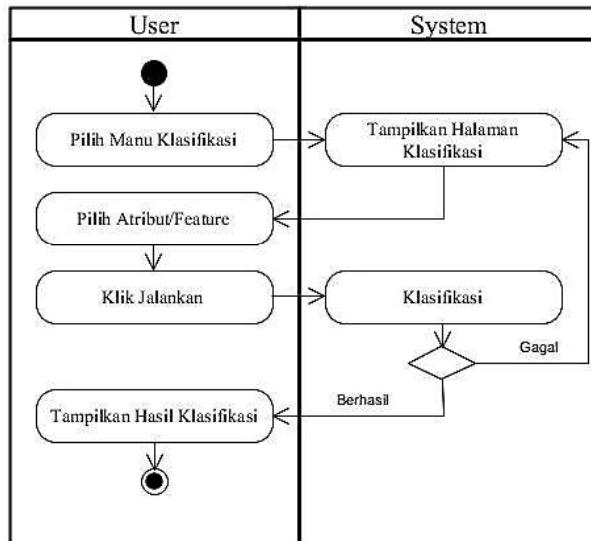
Activity Diagram input data latih untuk sistem yang akan dibuat pada penelitian ini dapat dilihat pada gambar 4.9.



Gambar 4.9 Activity Diagram Hapus Data Uji

5. Activity Diagram Klasifikasi

Activity Diagram input data latih untuk sistem yang akan dibuat pada penelitian ini dapat dilihat pada gambar 4.10.



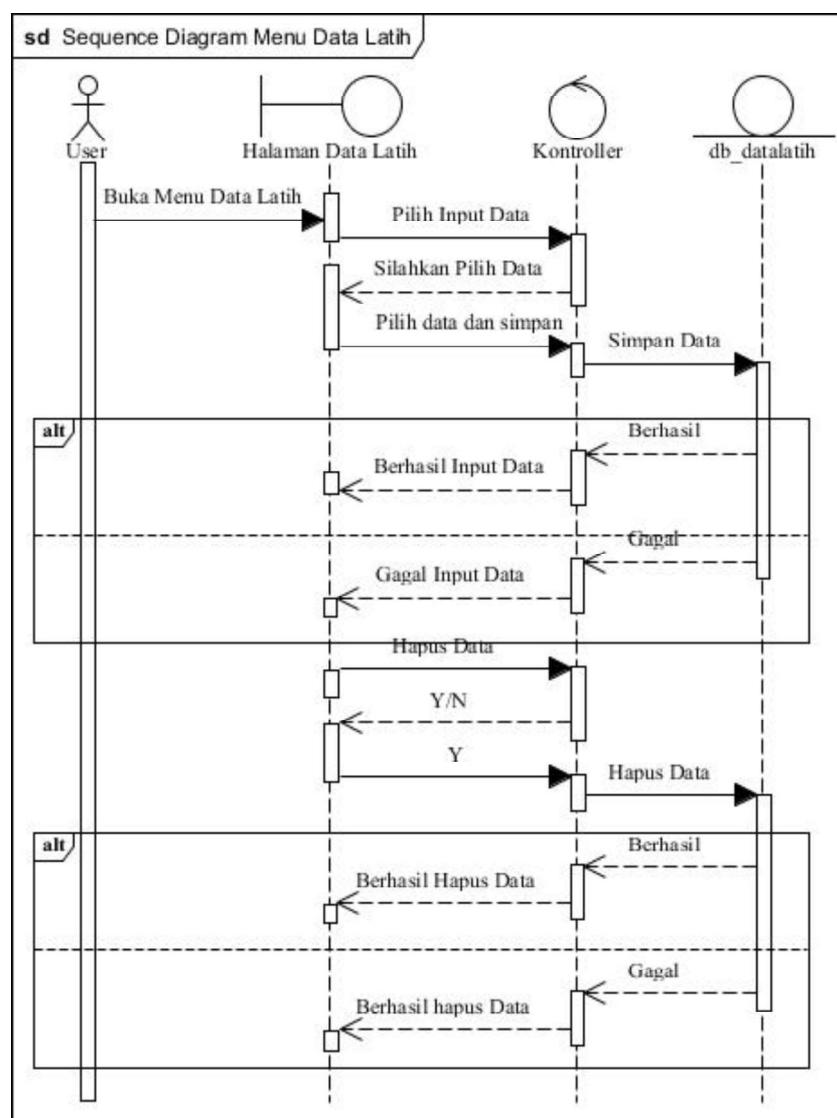
Gambar 4.10 Activity Diagram Klasifikasi

4.2.1.4 Sequence Diagram

Sequence diagram merupakan diagram yang menggambarkan kolaborasi dinamis antara sejumlah object, kegunaan *Sequence diagram* adalah untuk menunjukkan rangkaian pesan yang dikirim antara object juga interaksi antara object.

1. Sequence Diagram Menu Data Latih

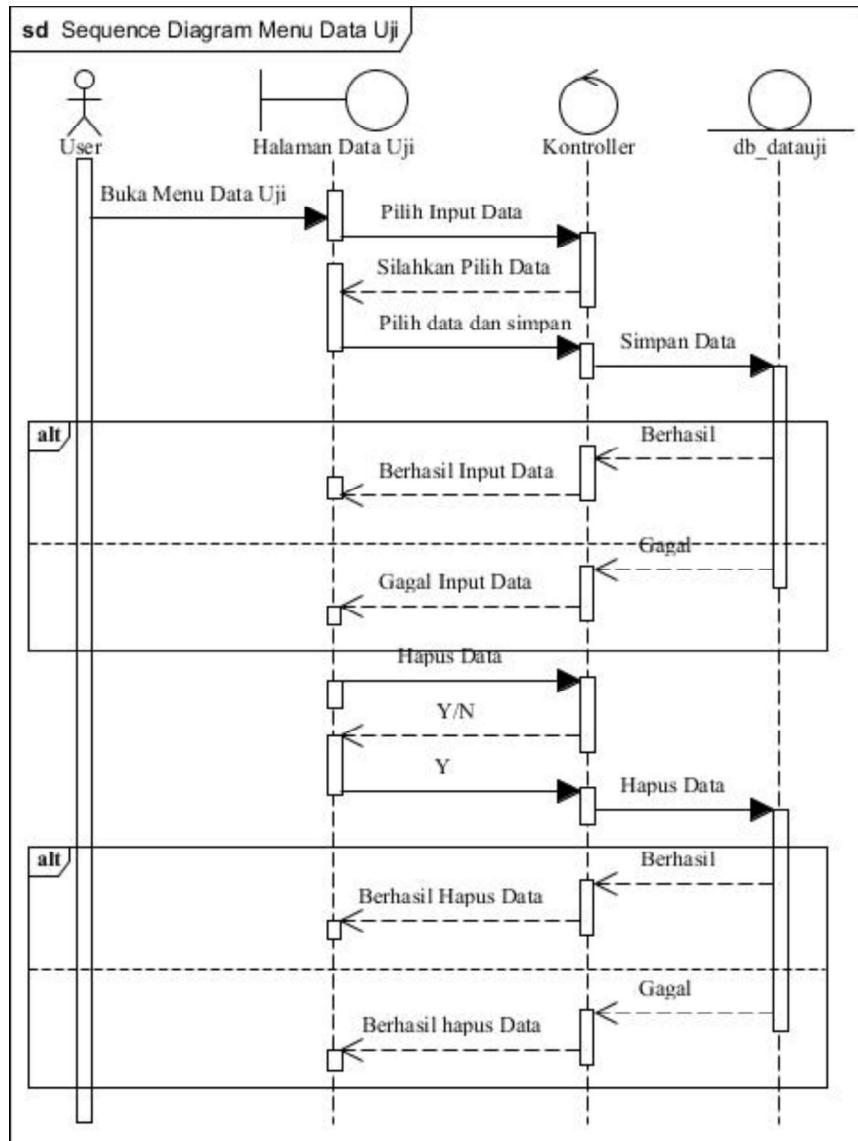
Sequence diagram menu data latih pada sistem yang akan dibuat pada penelitian kali ini dapat dilihat pada Gambar 4.11.



Gambar 4.11 Sequence Diagram Menu Data Latih

2. Sequence Diagram Menu Data Uji

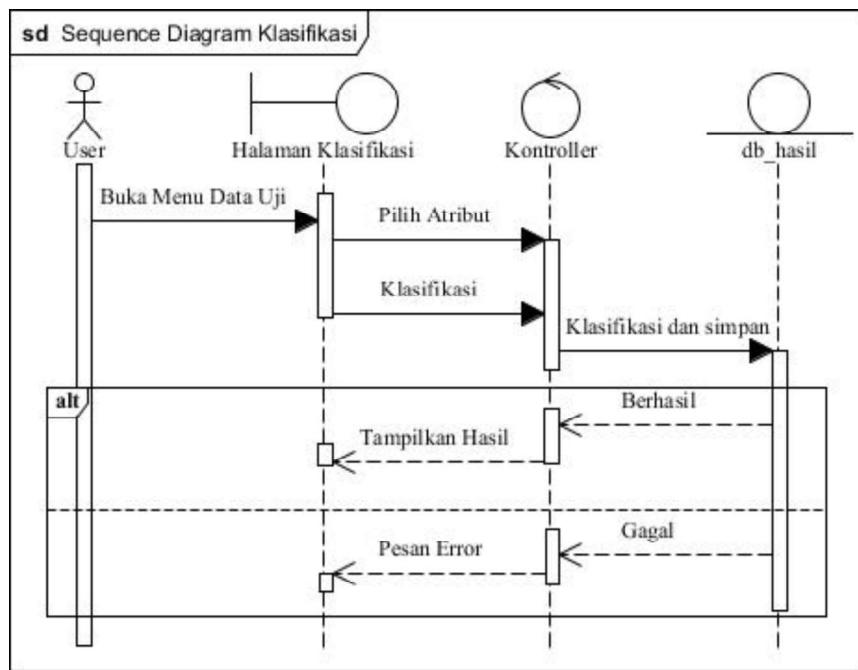
Sequence diagram menu data uji pada sistem yang akan dibuat pada penelitian ini dapat dilihat pada Gambar 4.12.



Gambar 4.12 Sequence Diagram Menu Data Uji

3. Sequence Diagram Klasifikasi

Sequence diagram menu klasifikasi pada sistem yang akan dibuat pada penelitian ini dapat dilihat pada Gambar 4.13.



Gambar 4.13 Sequence Diagram Klasifikasi

4.2.1.5 Class Diagram

Class diagram adalah model statis yang menggambarkan struktur dan deskripsi class serta hubungannya antara class, *Class diagram* untuk sistem yang akan dibuat pada penelitian ini dapat dilihat pada Gambar 4.14.

DataTraining	DataUji	Hasil
<pre>+Id_data_lath +duration +protocol_type +service +flag +src_bytes +dst_bytes +land +wrong_fragment +urgent +hot +num_failed_logins +logged_in +num_compromised +root_shell +su_attempted +num_root +num_file_creations +num_shells +num_access_fles +num_outbound_cmds +is_host_login +is_guest_login +Nama Field +count +srv_count +serror_rate +srv_serror_rate +error_rate +srv_error_rate +same_srv_rate +diff_srv_rate +srv_diff_host_rate +dst_host_count +dst_host_srv_count +dst_host_same_srv_rate +dst_host_diff_srv_rate +dst_host_same_src_port_rate +dst_host_srv_diff_host_rate +dst_host_error_rate +dst_host_srv_error_rate +dst_host_srv_error_rate +Class ----- +Import() +Delete()</pre>	<pre>+Id_data_lath +duration +protocol_type +service +flag +src_bytes +dst_bytes +land +wrong_fragment +urgent +hot +num_failed_logins +logged_in +num_compromised +root_shell +su_attempted +num_root +num_file_creations +num_shells +num_access_fles +num_outbound_cmds +is_host_login +is_guest_login +Nama Field +count +srv_count +serror_rate +srv_serror_rate +error_rate +srv_error_rate +same_srv_rate +diff_srv_rate +srv_diff_host_rate +dst_host_count +dst_host_srv_count +dst_host_same_srv_rate +dst_host_diff_srv_rate +dst_host_same_src_port_rate +dst_host_srv_diff_host_rate +dst_host_error_rate +dst_host_srv_error_rate +dst_host_srv_error_rate +Class ----- +Import() +Delete()</pre>	<pre>-Id -columns -atribut -normal -dos -r2l -probe -u2r ----- -Insert -Delete</pre>

Gambar 4.5 Class Diagram

4.2.2 Perancangan Database

Perancangan Table merupakan suatu perancangan tabel yang akan digunakan untuk menyimpan data pada suatu penyimpanan yang akan digunakan pada sistem.

4.2.2.1 Tabel Data Latih

Nama : `tbl_data_latih`

Deskripsi : Berisi data untuk pembelajaran klasifikasi

Primary key : `id`

Penjelasan struktur tabel `data_latih` dapat dilihat selengkapnya pada Tabel 4.27

Tabel 4.27 Prerancangan Tabel Data Latih

No	Nama Field	Tipe Dan Panjang Data	Null	Keterangan
1	<code>Id_data_latih</code>	<code>Int (11)</code>	<code>Not Null</code>	Primary key
2	<code>duration</code>	<code>Varchar (30)</code>	<code>Not Null</code>	
3	<code>protocol_type</code>	<code>Varchar (30)</code>	<code>Not Null</code>	
4	<code>service</code>	<code>Varchar (30)</code>	<code>Not Null</code>	
5	<code>flag</code>	<code>Varchar (30)</code>	<code>Not Null</code>	
6	<code>src_bytes</code>	<code>Varchar (30)</code>	<code>Not Null</code>	
7	<code>dst_bytes</code>	<code>Varchar (30)</code>	<code>Not Null</code>	
8	<code>land</code>	<code>Varchar (30)</code>	<code>Not Null</code>	
9	<code>wrong_fragment</code>	<code>Varchar (30)</code>	<code>Not Null</code>	
10	<code>urgent</code>	<code>Varchar (30)</code>	<code>Not Null</code>	
11	<code>hot</code>	<code>Varchar (30)</code>	<code>Not Null</code>	
12	<code>num_failed_logins</code>	<code>Varchar (30)</code>	<code>Not Null</code>	
13	<code>logged_in</code>	<code>Varchar (30)</code>	<code>Not Null</code>	
14	<code>num_compromised</code>	<code>Varchar (30)</code>	<code>Not Null</code>	
15	<code>root_shell</code>	<code>Varchar (30)</code>	<code>Not Null</code>	
16	<code>su_attempted</code>	<code>Varchar (30)</code>	<code>Not Null</code>	
17	<code>num_root</code>	<code>Varchar (30)</code>	<code>Not Null</code>	
18	<code>num_file_creations</code>	<code>Varchar (30)</code>	<code>Not Null</code>	
19	<code>num_shells</code>	<code>Varchar (30)</code>	<code>Not Null</code>	
20	<code>num_access_files</code>	<code>Varchar (30)</code>	<code>Not Null</code>	
21	<code>num_outbound_cmds</code>	<code>Varchar (30)</code>	<code>Not Null</code>	
22	<code>is_host_login</code>	<code>Varchar (30)</code>	<code>Not Null</code>	
23	<code>is_guest_login</code>	<code>Varchar (30)</code>	<code>Not Null</code>	

No	Nama Field	Tipe Dan Panjang Data	Null	Keterangan
24	count	Varchar (30)	Not Null	
25	srv_count	Varchar (30)	Not Null	
26	serror_rate	Varchar (30)	Not Null	
27	srv_serror_rate	Varchar (30)	Not Null	
28	error_rate	Varchar (30)	Not Null	
29	srv_error_rate	Varchar (30)	Not Null	
30	same_srv_rate	Varchar (30)	Not Null	
31	diff_srv_rate	Varchar (30)	Not Null	
32	srv_diff_host_rate	Varchar (30)	Not Null	
33	dst_host_count	Varchar (30)	Not Null	
34	dst_host_srv_count	Varchar (30)	Not Null	
35	dst_host_same_srv_rate	Varchar (30)	Not Null	
36	dst_host_diff_srv_rate	Varchar (30)	Not Null	
37	dst_host_same_src_port_rate	Varchar (30)	Not Null	
38	dst_host_srv_diff_host_rate	Varchar (30)	Not Null	
39	dst_host_serror_rate	Varchar (30)	Not Null	
40	dst_host_srv_serror_rate	Varchar (30)	Not Null	
41	dst_host_error_rate	Varchar (30)	Not Null	
42	dst_host_srv_error_rate	Varchar (30)	Not Null	
43	Class	Varchar (100)	Not Null	

4.2.2.2 Tabel Data Uji

Nama : tbl_data_uji

Deskripsi : Berisi data untuk pengujian klasifikasi

Primary key : id

Penjelasan struktur tabel data_uji dapat dilihat selengkapnya pada Tabel 4.28.

Tabel 4.28 Perancangan Tabel Data Uji

No	Nama Field	Tipe Dan Panjang Data	Null	Keterangan
1	id	Int (11)	Not Null	Primary key
2	duration	Varchar (30)	Not Null	
3	protocol_type	Varchar (30)	Not Null	
4	service	Varchar (30)	Not Null	

No	Nama Field	Tipe Dan Panjang Data	Null	Keterangan
5	flag	Varchar (30)	Not Null	
6	src_bytes	Varchar (30)	Not Null	
7	dst_bytes	Varchar (30)	Not Null	
8	land	Varchar (30)	Not Null	
9	wrong_fragment	Varchar (30)	Not Null	
10	urgent	Varchar (30)	Not Null	
11	hot	Varchar (30)	Not Null	
12	num_failed_logins	Varchar (30)	Not Null	
13	logged_in	Varchar (30)	Not Null	
14	num_compromised	Varchar (30)	Not Null	
15	root_shell	Varchar (30)	Not Null	
16	su_attempted	Varchar (30)	Not Null	
17	num_root	Varchar (30)	Not Null	
18	num_file_creations	Varchar (30)	Not Null	
19	num_shells	Varchar (30)	Not Null	
20	num_access_files	Varchar (30)	Not Null	
21	num_outbound_cmds	Varchar (30)	Not Null	
22	is_host_login	Varchar (30)	Not Null	
23	is_guest_login	Varchar (30)	Not Null	
24	count	Varchar (30)	Not Null	
25	srv_count	Varchar (30)	Not Null	
26	serror_rate	Varchar (30)	Not Null	
27	srv_serror_rate	Varchar (30)	Not Null	
28	rerror_rate	Varchar (30)	Not Null	
29	srv_rerror_rate	Varchar (30)	Not Null	
30	same_srv_rate	Varchar (30)	Not Null	
31	diff_srv_rate	Varchar (30)	Not Null	
32	srv_diff_host_rate	Varchar (30)	Not Null	
33	dst_host_count	Varchar (30)	Not Null	
34	dst_host_srv_count	Varchar (30)	Not Null	
35	dst_host_same_srv_rate	Varchar (30)	Not Null	
36	dst_host_diff_srv_rate	Varchar (30)	Not Null	
37	dst_host_same_src_port_rate	Varchar (30)	Not Null	
38	dst_host_srv_diff_host_rate	Varchar (30)	Not Null	
39	dst_host_serror_rate	Varchar (30)	Not Null	
40	dst_host_srv_serror_rate	Varchar (30)	Not Null	

No	Nama Field	Tipe Dan Panjang Data	Null	Keterangan
41	dst_host_error_rate	Varchar (30)	Not Null	
42	dst_host_srv_error_rate	Varchar (30)	Not Null	
43	Class	Varchar (100)	Not Null	

4.2.2.3 Tabel Hasil

Nama : hasil

Deskripsi : Berisi data hasil dari klasifikasi

Primary key : id_hasil

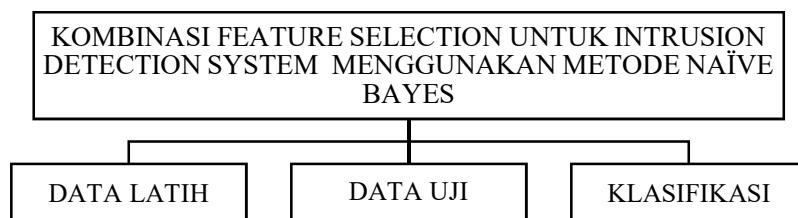
Penjelasan struktur tabel data_uji dapat dilihat selengkapnya pada Tabel 4.29.

Tabel 4.29 Perancangan Tabel Hasil

No	Nama Field	Tipe Dan Panjang Data	Null	Keterangan
1	Id	Int (11)	Not Null	Primary key
2	columns	Varchar (30)	Not Null	
3	atribut	Varchar (30)	Not Null	
4	normal	Varchar (30)	Not Null	
5	dos	Varchar (30)	Not Null	
6	r2l	Varchar (30)	Not Null	
7	probe	Varchar (30)	Not Null	
8	u2r	Varchar (30)	Not Null	

4.2.3 Perancangan struktur menu

Perancangan struktur menu merupakan perancangan menu apa saja yang akan digunakan pada sebuah sistem yang sesuai dengan kebutuhan. Berikut merupakan struktur menu yang dapat dilihat ada gambar 4.15.



Gambar 4.14 Perancangan Struktur Menu

Berdasarkan gambar diatas, terdapat 3 menu yang akan digunakan pada sistem ini.

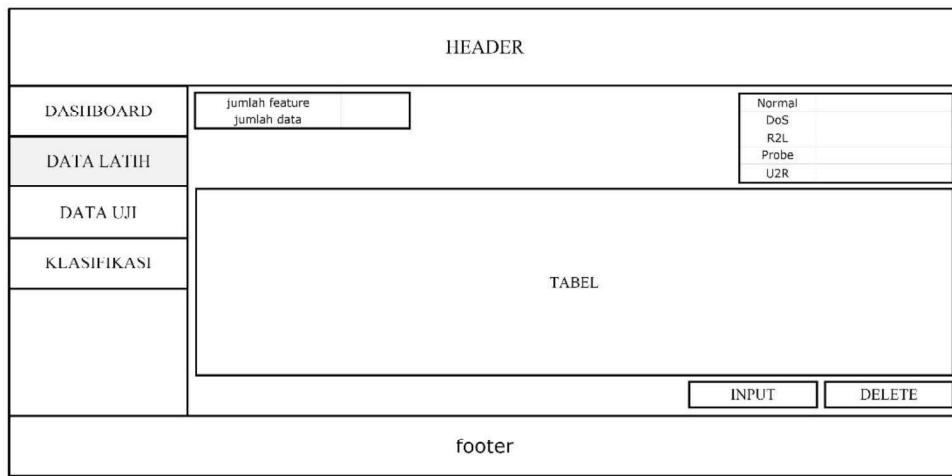
1. Menu Data Latih, yaitu menu untuk menampilkan data latih yang akan kita gunakan pada sistem ini, selain menampilkan data latih pada menu ini juga bisa untuk menambah dan menghapus data latih oleh usernya.
2. Menu Data Uji, yaitu yaitu menu untuk menampilkan data Uji yang akan kita gunakan pada sistem ini, selain menampilkan data uji pada menu ini juga bisa untuk menambah dan menghapus data uji oleh usernya.
3. Menu Klasifikasi, yaitu menu untuk melakukan klasifikasi menggunakan metode klasifikasi Naïve Bayes, dan menampilkan hasil dari klasifikasi.

4.2.4 Perancangan Antar Muka (*Interface*)

Antarmuka (*interface*) sistem merupakan sarana pengembangan sistem yang akan digunakan untuk membuat komunikasi dan penyampaian informasi lebih mudah dimengerti, serta konsisten antara sistem dengan user. Interface meliputi tampilan yang mudah dimengerti, baik serta memiliki tombol – tombol yang familiar dan user friendly.

4.3.4.1 Perancangan halaman data latih

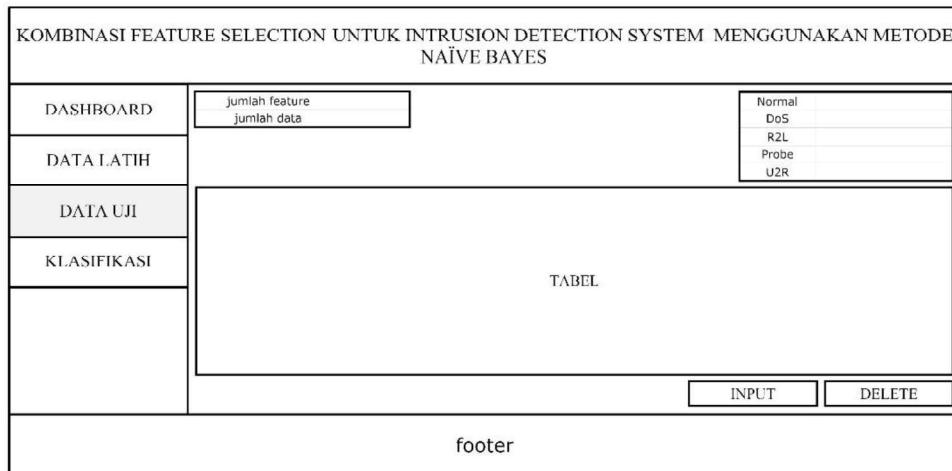
Halaman data latih merupakan halaman untuk menampilkan data latih yang akan digunakan pada penelitian, selain menampilkan data latih pada halaman ini user juga bisa melakukan tambah dan hapus data latih. Halaman data latih dapat dilihat pada Gambar 4.16.



Gambar 4.15 Perancangan Halaman Data Latih

4.3.4.2 Perancangan halaman data uji

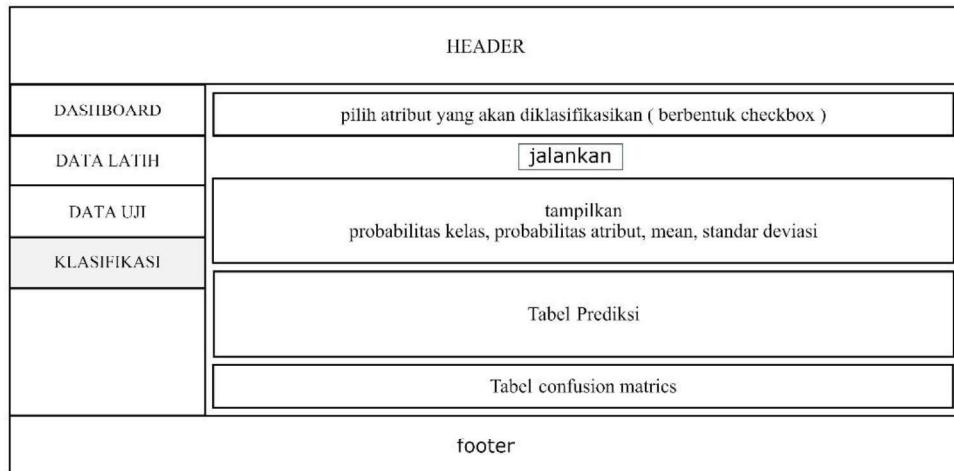
Halaman data uji merupakan halaman untuk menampilkan data uji yang akan digunakan pada penelitian ini, selain menampilkan data uji pada halaman ini user juga bisa melakukan tambah dan hapus data uji tersebut. Halaman data uji dapat dilihat pada Gambar 4.17



Gambar 4.16 Perancangan Halaman Data Uji

4.3.4.3 Perancangan halaman klasifikasi

Menu Klasifikasi, yaitu menu untuk melakukan klasifikasi menggunakan metode klasifikasi Naïve Bayes, dan menampilkan hasil dari klasifikasi. Rancangan halaman Klasifikasi dapat dilihat pada Gambar 4.18.



Gambar 4.17 Perancangan halaman Klasifikasi