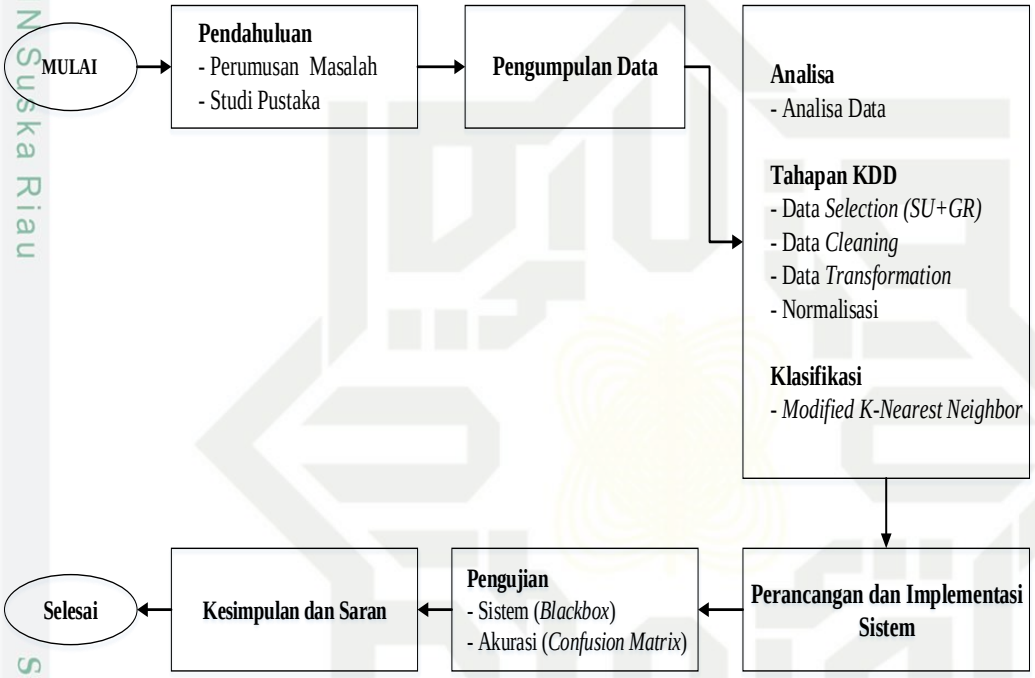
BAB III
METODOLOGI PENELITIAN

Dalam metodologi penelitian terdapat beberapa langkah sistematis yang digunakan sebagai acuan dalam melaksanakan penelitian. Gambaran umum tentang langkah dalam penelitian tugas akhir ini adalah sebagai berikut:



Gambar 3.1 Tahapan Dalam Metodologi Penelitian

3.1 Pendahuluan

Tahapan ini merupakan persiapan awal untuk memulai penelitian yang akan dilakukan. Persiapan awal ini menjadi sebuah acuan untuk melakukan penelitian. Terdapat dua langkah persiapan yang dilakukan, diantaranya adalah:

3.1.1 Perumusan Masalah

Pada tahapan ini dilakukan pembelajaran terhadap masalah yang terjadi di dalam penelitian ini dan merupakan tahapan awal dalam memulai sebuah penelitian. Berdasarkan permsalahan yang didapati kemudian dilakukan sebuah solusi untuk memecahkan masalah tersebut. Permasalahan pada penelitian kali ini adalah bagaimana menerapkan kombinasi *feature selection Gain Ratio* dan

Symmetrical Uncertainly dengan metode klasifikasi *Modified K-Nearest Neighbor* (MK-NN) menggunakan *dataset* NSL-KDD.

3.1.2 Studi Pustaka

Tahapan selanjutnya setelah merumuskan masalah adalah melakukan pengumpulan informasi melalui penelitian terdahulu dan buku-buku yang bersangkutan dengan klasifikasi menggunakan metode pada *data mining*.

3.2 Pengumpulan Data

Data yang digunakan pada penelitian ini adalah *dataset* NSL-KDD. *Dataset* yang digunakan pada penelitian ini diperoleh dari situs resmi *github.com* yang dibagikan secara umum melalui https://github.com/defcom17/NSL_KDD pada tahun 2015. Jumlah *record* data dari dataset ini sebanyak 125.973 yang terdiri dari 5 kelas dan 41 parameter. Terdapat 2 jumlah data yang akan digunakan dalam penelitian ini yaitu data yang berjumlah 1052 (tidak setara) dan data yang berjumlah 25 (setara) untuk dilakukan pengujian.

3.3 Analisa

Tahap ini akan dilakukan analisa terhadap beberapa proses dari penelitian yang akan dilakukan. Adapun proses yang dilakukan pada penelitian ini sebagai berikut:

3.3.3 Analisa Data

Di dalam *dataset* NSL KDD terdapat 41 parameter yang akan digunakan pada penelitian ini, berikut akan ditampilkan pada Tabel 3.1

Tabel 3.1 Parameter Dataset NSL-KDD

No	Parameter	Keterangan	*Type*
1	*Duration* (F1)	Lama waktu koneksi	*Numeric*
2	*Protocol_type* (F2)	Tipe *protocol* yang digunakan pada koneksi	*Nominal*
3	*Service* (F3)	Tujuan *Network Service* yang digunakan	*Nominal*
4	*Flag* (F4)	Status koneksi (normal atau *error*)	*Nominal*
5	*Src_bytes* (F5)	Jumlah *bytes* yang di transfer dari sumber ke tujuan dalam satu koneksi	*Numeric*

No	Parameter	Keterangan	*Type*
6	*Dst_bytes* (F6)	Jumlah *bytes* yang ditransfer dari tujuan ke sumber dalam satu koneksi	*Numeric*
7	*Land* (F7)	Jika sumber dan tujuan alamat IP dan nomor port sama maka variabel ini akan bernilai 1 atau 0	*Binary*
8	*Wrong_fragment* (F8)	Jumlah *Wrong fragment* pada koneksi	*Numeric*
9	*Urgent* (F9)	Jumlah paket yang *urgent* pada koneksi	*Numeric*
10	*Hot* (F10)	Jumlah indikator '*Hot*' pada *content* seperti: mengakses sistem direktori, membuat *programs* dan menjalankan *programs*	*Numeric*
11	*Num_failed_logins* (F11)	Jumlah percobaan login yang gagal	*Numeric*
12	*Logged_in* (F12)	Status login (bernilai 1 jika login sukses dan bernilai 0 jika gagal)	*Binary*
13	*Num_compromised* (F13)	Jumlah kondisi *compromised*	*Numeric*
14	*Root_shell* (F14)	Bernilai 1 jika *root shell* didapat dan bernilai 0 jika tidak	*Binary*
15	*Su_attempted* (F15)	Bernilai 1 jika dilakukan percobaan atau menjalankan perintah '*su root*' dan bernilai 0 apabila tidak dijalankan	*Binary*
16	*Num_root* (F16)	Jumlah *root* yang diakses atau jumlah operasi yang diakses sebagai *root* pada koneksi	*Numeric*
17	*Num_file_creations* (F17)	Jumlah operasi *file creations* pada koneksi	*Numeric*
18	*Num_shells* (F18)	Jumlah *shell prompts*	*Numeric*
19	*Num_access_files* (F19)	Jumlah operasi pada akses *file control*	*Numeric*
20	*Num_outbound_cmds* (F20)	Jumlah *outbound command* pada *ftp session*	*Numeric*
21	*Is_hot_login* (F21)	Bernilai 1 jika login yang terdapat pada *list hot* (*root* atau admin)	*Binary*
22	*Is_guest_login* (F22)	Bernilai 1 jika login sebagai '*guest login*' dan sebaliknya	*Binary*
23	*Count* (F23)	Jumlah koneksi yang mengarah ke tujuan *host* yang sama pada jaringan yang sedang berjalan	*Numeric*
24	*Srv_count* (F24)	Jumlah koneksi dengan *service* yang sama (*port number*) pada koneksi yang sedang berjalan	*Numeric*

No	Parameter	Keterangan	*Type*
25	*Serror_rate* (F25)	Persentase koneksi yang diaktifkan *flag* (4) s0, s1, s2 atau s3 diantara *connection aggregated* pada *count* (23)	*Numeric*
26	*Srv_serror_rate* (F26)	Persentase koneksi yang diaktifkan *flag* (4) s0, s1, s2, atau s3 diantara *connection aggregated* pada *srv_count* (24)	*Numeric*
27	*Rerror_rate* (F27)	Persentase koneksi yang diaktifkan *flag* (4) REJ, diantara *connection aggregated* pada *count* (23)	*Numeric*
28	*Srv_rerror_rate* (F28)	Persentase koneksi yang diaktifkan *flag* (4) REJ, diantara *connection aggregated* pada *srv_count* (24)	*Numeric*
29	*Same_srv_rate* (F29)	Persentase koneksi pada *service* yang sama, diantara *connection aggregated* pada *count* (23)	*Numeric*
30	*Diff_srv_rate* (F30)	Persentase koneksi pada *service* yang berbeda, diantara *connection aggregated* pada *count* (23)	*Numeric*
31	*Srv_diff_host_rate* (F31)	Persentase koneksi ke tujuan *machines* yang berbeda diantara *connection aggregated* pada *srv_count* (24)	*Numeric*
32	*Dst_host_count* (F32)	Jumlah koneksi yang memiliki *host op address* tujuan yang sama	*Numeric*
33	*Dst_host_srv_count* (F33)	Jumlah koneksi yang memiliki *port number* yang sama	*Numeric*
34	*Dst_host_same_srv_rate* (F34)	Persentase koneksi dengan *service* yang sama diantara *connections aggregated* pada *dst_host_count* (32)	*Numeric*
35	*Dst_host_diff_srv_rate* (F35)	Persentase koneksi dengan *service* yang berbeda, diantara *connections aggregated* pada *dst_host_count* (32)	*Numeric*
36	*Dst_host_same_src_port_rate* (F36)	Persentase koneksi dengan sumber port yang sama diantara connections aggregated pada dst_host_srv_count (33)	*Numeric*
37	*Dst_host_srv_diff_host_rate* (F37)	Persentase koneksi dengan tujuan *machines* yang berbeda, diantara *connections aggregated* pada *dst_host_srv_count* (33)	*Numeric*
38	*Dst_host_serror_rate* (F38)	Persentase koneksi yang diaktifkan *flag* (4) s0, s1, s2 atau s3 diantara	*Numeric*

No	Parameter	Keterangan	*Type*
		connectons aggregated pada *dst_host_count* (32)	
39	*Dst_host_srv_serror_rate* (F39)	Persentase koneksi yang diaktifkan *flag* (4) s0, s1 atau s3 diantara *connections aggregated* pada *dst_host_srv_count* (33)	*Numeric*
40	*Dst_host_rerror_rate* (F40)	Persentase koneksi yang diaktifkan *flag* (4) REJ, diantara *connections aggregated* pada *dst_host_count* (32)	*Numeric*
41	*Dst_host_srv_rerror_rate* (F41)	Persentase koneksi yang diaktifkan *flag* (4) REJ, diantara *connections aggregated* pada *dst_host_srv_count* (33)	*Numeric*

Jumlah *dataset* NSL-KDD dikelompokkan menjadi lima kategori serangan, kategori serangan dapat dilihat pada Tabel 3.2 berikut:

Tabel 3.2 Kategori Serangan *Dataset* NSL-KDD

Jenis Serangan	Jumlah Data
Normal	67.343
DoS	45.927
ProbeS	11.656
U2R	52
R2L	995
Total	**125.973**

Daftar kelas serangan yang digunakan pada *dataset* NSL-KDD dapat dilihat pada Tabel 3.3 berikut:

Tabel 3.3 Daftar Kelas Serangan *Dataset* NSL KDD

No	Kelas Serangan	Tipe Serangan
1	*DoS*	*Back, Land, Neptune, Pod, Smurf, Teardrop, Apache2, Udpstorm, Processtable, Worm*
2	*ProbeS*	*Satan, Ipsweep, Nmap, Portsweep, Mscan, Saint*
3	*R2L*	*Guess_Password, Ftp_write, Imap, Phf, Multihop, Warezmaster, Warezclient, Spy, Xlock, Xsnoop, Snmpguess, Snmpgetattack, Httptunnel, Sendmail, Named*
4	*U2R*	*Buffer_overflow, Loadmodule, Rootkit, Perl, Sqlattack, Xterm, Ps*

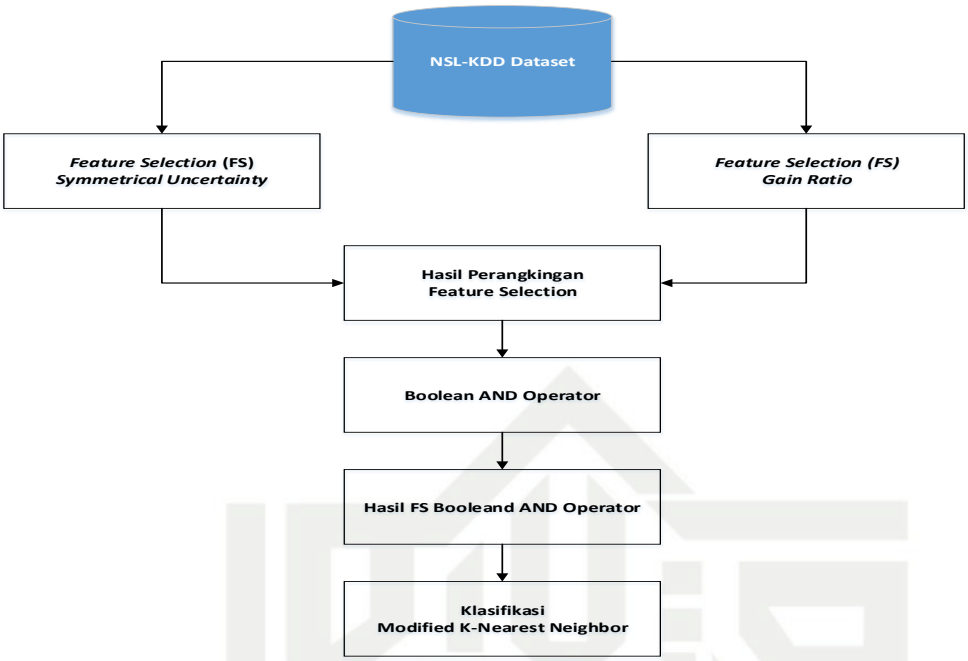
3.3.4 Tahapan *Knowledge Discovery in Database* (KDD)

Tahapan *Knowledge Discovery in Database* (KDD) merupakan tahapan yang menjelaskan beberapa proses yang dilakukan dalam *data mining* hingga melakukan klasifikasi menggunakan metode *Modified K-Nearest Neighbor* (MK-NN). Beberapa tahapan yang akan dilakukan seperti *data selection*, *data celaning* (*pre-processing*), *data transformation* hingga klasifikasi menggunakan metode MK-NN.

3.3.4.1 *Data Selection*

Tahapan ini dilakukannya proses pemilihan (seleksi) data yang berperan penting dalam menentukan jenis serangan menggunakan metode *feature selection* (FS). Pada penelitian ini metode FS yang digunakan adalah kombinasi *symmetrical uncertainty* dan *gain ratio* yang dimulai dengan mencari *Information Gain* (IG) dan mencari *entropy* data (ID) serta *entropy* atribut (IA). Setelah mendapatkan nilai IG beserta nilai ID dan IA, selanjutnya menghitung *Symmetrical Uncertainly* (SU) hingga memperoleh hasil. Proses selanjutnya juga dilakukan pencarian nilai *gain ratio* (GR) hingga memperoleh hasil.

Hasil dari tahapan ini yaitu diperolehnya nilai bobot pada masing-masing atribut atau fitur dan kemudian dilakukan perangkingan berdasarkan nilai tertinggi ke terendah. Dalam menentukan fitur yang akan digunakan untuk klasifikasi, pada penelitian ini diambil 20 nilai tertinggi yang dilakukan berdasarkan penelitian yang telah dilakukan sebelumnya oleh (Garg and Kumar, 2014) dengan judul *Combinational Feature Selection Approach for Network Intrusion Detection System*. Pada penelitian tersebut dikatakan bahwa pemilihan fitur untuk klasifikasi menggunakan kombinasi *feature selection* dapat dilakukan dengan mengambil 15 hingga 20 fitur teratas. Setelah mendapatkan 20 fitur teratas dari kedua FS, selanjutnya dilakukan proses *boolean AND operator* yaitu dengan mencari irisan atau keterkaitan antara fitur yang diperoleh dari SU dan GR, fitur yang diperoleh tersebut kemudian digunakan untuk klasifikasi. Alur proses *feature selection* dapat dilihat pada Gambar 3.2 Berikut:



Gambar 3.2 Proses *Feature Selection*

3.3.4.2 *Data Cleaning*

Pada tahapan ini dilakukan proses pembersihan atau penghapusan terhadap data, seperti data *missing value*, duplikasi data, data yang inkonsisten. Pada dataset NSL-KDD ini tidak terdapat data yang bersifat *missing value*, duplikasi data, maupun data yang inkonsisten.

3.3.4.3 *Data Transformation*

Tahapan ini dilakukannya proses merubah nilai atribut yang berbentuk simbol atau huruf menjadi angka. Pada dataset NSL-KDD yang digunakan pada penelitian ini, dari 41 atribut terdapat 3 atribut yang bernilai huruf dan akan ditransformasi, yaitu *protocol_type*, *service*, *flag*. Kemudian merubah nilai-nilai angka tersebut menjadi berkisar antara 0-1 (normalisasi) menggunakan Persamaan (2.14). Contohnya akan ditampilkan pada Tabel 3.4 Berikut:

Tabel 3.4 Transformasi *Protocol_type*

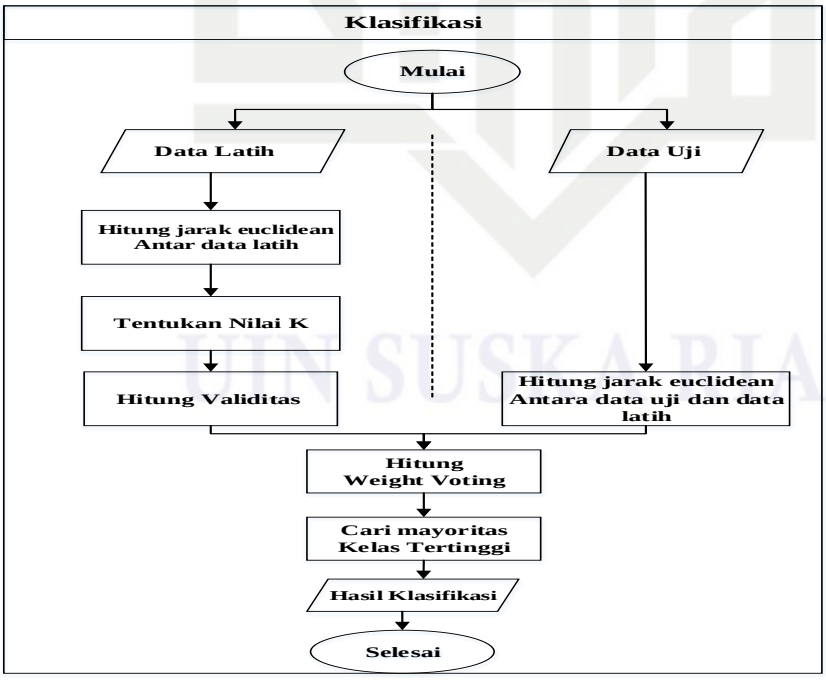
Protocol_type	No
TCP	1
UDP	2
ICMP	3

3.3.4.4 Klasifikasi

Tahapan ini dilakukan setelah tahap *preprocessing* selesai, tahap ini menunjukkan proses klasifikasi kategori serangan jaringan yang terdapat pada *dataset* NSL-KDD menggunakan metode *Modified K-Nearest Neighbor* (KN-N) dengan mengolah data yang ada. Metode MK-NN melakukan klasifikasi kategori serangan berdasarkan data latih yang mempunyai jarak tetangga terdekat. Adapun tahapan di dalam metode MK-NN pada penelitian ini adalah sebagai berikut:

a. Menentukan nilai K
b. Menghitung jarak antara data uji terhadap setiap data latih yang ada menggunakan rumus *euclidean* (Persamaan 2.8).
c. Meghitung validitas dari data latih yang ada (Persamaan 2.9).
d. Menghitung *weight voting* (Persamaan 2.10).
e. Menentukan kelas mayoritas dari K buah data latih dengan *weight voting* yang tertinggi.
f. *Output* yang dihasilkan merupakan hasil klasifikasi kelas kategori serangan jaringan dari dataset NSL-KDD berupa kelas data *Normal*, *Dos*, *ProbeS*, U2R, R2L.

Berikut merupakan *flowchart* dari metode *Modified K-Nearest Neighbor* yang dapat dilihat pada Gambar 3.3:



Gambar 3.3 *Flowchart* Metode *Modified K-Nearest Neighbor*

3.4 Perancangan dan Implementasi Sistem

Perancangan dan implementasi sistem dilakukan setelah model klasifikasi menggunakan metode MK-NN selesai, sistem akan dibangun berdasarkan perancangan seperti berikut:

a. Perancangan *Database*

Perancangan *database* dilakukan dengan membuat sebuah desain *database* dalam bentuk konseptual model yang akan digunakan untuk penyimpanan data.

b. Perancangan Struktur Menu

Perancangan struktur menu dilakukan dengan merancang menu-menu pada sistem yang sesuai dengan fungsinya masing-masing.

c.* Perancangan *Interface

Dilakukan perancangan antarmuka sistem yang membuat interaksi antar pengguna dengan sistem. Tampilan yang dibuat menghasilkan gambaran umum implementasi dari sistem yang dibuat.

3.5 Implementasi Sistem

Implementasi sistem dikembangkan berdasarkan analisa dan perancangan yang telah dilakukan sebelumnya yang bertujuan agar sistem dapat dioperasikan pada keadaan yang sebenarnya sehingga dapat diketahui apakah hasil yang diperoleh dari sistem sesuai dengan tujuan penelitian yang dilakukan. Adapun perangkat pendukung yang dibutuhkan adalah sebagai berikut:

Perangkat keras (*Hardware*)

Processor	: Intel *Core* i5-8259U CPU @3.20 Ghz
RAM	: 4 GB

Perangkat lunak (*Software*)

Sistem Operasi	: *Windows* 10
Bahasa Pemrograman	: PHP (*Object Oriented Programming*).
Tools	: *Visual Studio Code, Xampp Control Panel.*
Browser	: *Google Chrome*

3.6 Pengujian

Setelah melakukan implementasi sistem, selanjutnya dilakukan evaluasi terhadap sistem yang telah dibuat. Evaluasi dilakukan dengan proses pengujian untuk memastikan apakah sistem yang dibangun sesuai dengan hasil analisa dan perancangan, serta mengetahui tingkat akurasi yang dihasilkan. Pengujian pada penelitian ini meliputi:

1. Pengujian *Blackbox.*
2. Pengujian Perhitungan Sistem.
3. Pengujian akurasi berdasarkan parameter MK-NN menggunakan Persamaan (2.11) dengan nilai K= 1, 3, 5, 7, 9, 11.

3.7 Kesimpulan dan Saran

Tahapan ini berisi rangkuman penelitian dan kesimpulan yang sesuai dengan rumusan masalah serta tujuan yang ingin dicapai. Pada tahapan ini juga berisi hal yang disarankan bagi pembaca untuk melakukan pengembangan yang berkaitan dengan penelitian ini di masa yang akan datang.