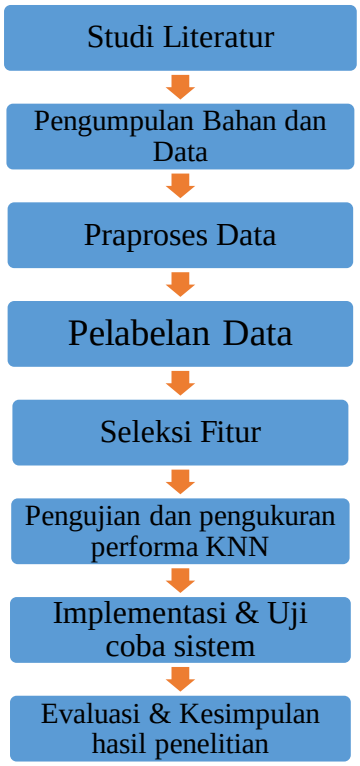
BAB III

METODOLOGI PENELITIAN

3.1 Tahapan Penelitian

Tahapan penelitian digambarkan secara lengkap dengan menggunakan *flow chart* dengan menggunakan metode penelitian kuantitatif. *Flow chart* atau diagram alur digunakan untuk memudahkan penyampaian informasi terkait langkah-langkah yang akan dilakukan dalam penelitian ini.Tahapan penelitian secara keseluruhan disajikan pada gambar 3.1.



Gambar 3.1 Tahapan Penelitian

3.1.1 Studi Literatur

Melakukan kajian pustaka yang menunjang penelitian diantaranya terkait analisis *malware*, *machine learning*¸ algoritma KNN dan teknik *information gain*. Literatur yang dikaji diperoleh dari jurnal – jurnal dan buku yang sumbernya terpercaya.

3.1.2 Pengumpulan Bahan dan Data

Data yang dibutuhkan untuk penelitian ini adalah data sekunder dari hasil penelitian Canadian Institut for Cybersecurity yaitu dataset CICDarknet 2020 yang di dalamnya berisi data data hasil tangkapan jaringan nyata yang memiliki trafik darknet diantaranya TOR dan VPN sebagai data latih.

Data yang digunakan diharuskan memiliki kualitas data yang baik, oleh karena itu diperlukan proses ekstraksi fitur, validasi, integrasi dan transformasi serta reduksi ukuran data dan diskretisasi data agar sesuai dengan data latih yang telah disiapkan (Rolansa, Yunita, dan Suheri 2020).

3.1.3 Pelabelan Data

Proses ekstraksi data akan menghasilkan data dengan format CSV yang memiliki banyak fitur yang sama seperti data latih, namun untuk fitur label tidak akan ada dan tidak dapat diklasifikasi secara otomatis. Pelabelan dilakukan secara manual dengan memperhatikan skema serangan yang dilakukan sebelumnya (Hafid 2019).

3.1.4 Seleksi Fitur

Seleksi fitur dimaksudkan untuk menyaring fitur mana saja yang berkaitan dengan label sehingga proses analisa dapat lebih cepat.

3.1.5 Pengujian dan Pengukuran Performa KNN

Hasil data yang telah diseleksi akan dilakukan pelatihan dan pengujian dengan menggunakan algoritma KNN untuk melihat nilai performa dari dataset yang digunakan. *Hypertuning parameter* atau pengujian nilai k secara berulang adalah salah satu teknik yang digunakan untuk melakukan pelatihan dataset untuk mendapatkan parameter k terbaik. K terbaik akan menjadi tolak ukur pengujian, percobaan dilakukan sebanyak dua kali, yaitu pelatihan dan pengujian dengan fitur lengkap serta dengan fitur terpilih.

3.1.6 Implementasi dan Uji Coba Sistem

Implementasi dilakukan dengan pembuatan sebuah sistem yang dapat memprediksi ancaman serangan yang terdapat dalam file PCAP dengan menggunakan algoritma KNN. Bahasa pemrograman yang digunakan adalah python yang memiliki modul untuk proses klasifikasi dan prediksi. Uji coba sistem akan dilakukan dengan menilai seberapa lama pemroresan klasifikasi dan prediksi data uji yang diunggah ke sistem.

3.1.7 Evaluasi & Kesimpulan Hasil Penelitian

Evaluasi dilaksanakan dengan menggunakan metode *Confussion Matrix* yang biasanya digunakan dalam evaluasi model pada kasus klasifikasi untuk menghitung tingkat akurasi, presisi, *recall* dan *f1-score.*

Hasil penelian akan disimpulkan dengan memerhatikan hasil pengukuran algoritma KKN dengan fitur lengkap dan fitur terpilih setelah dilakukan seleksi fitur.