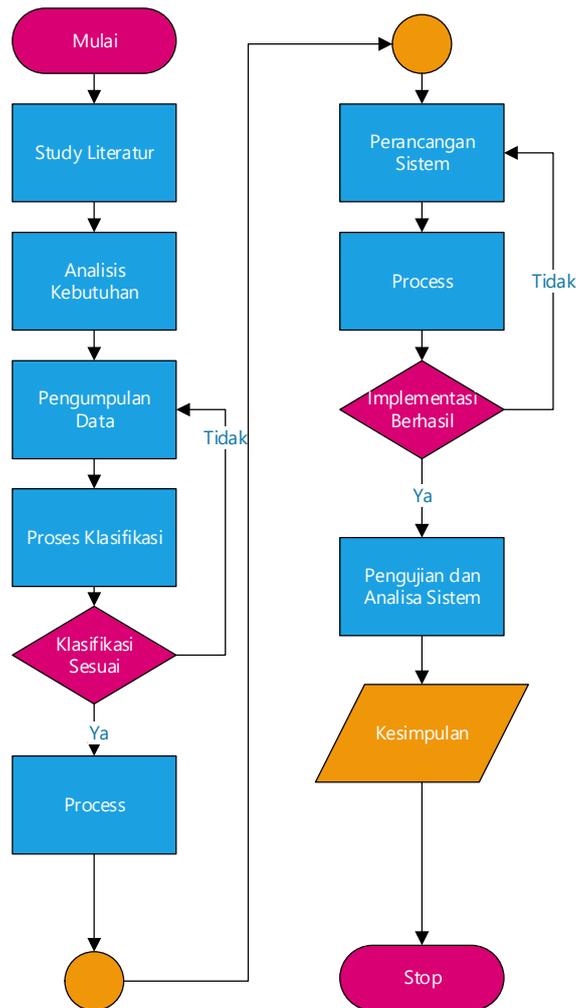


BAB 3 METODOLOGI

Pada bab ini pembahasan yang dibahas mengenai metode-metode yang dipergunakan untuk Implementasi Naive Byes Pada Proxy Server Untuk Klasifikasi Pengguna Internet. Beberapa hal yang dibahas dalam bab ini adalah Studi Literatur, Analisis Kebutuhan, Pengumpulan Data, Proses Klasifikasi, lalu dilanjutkan dengan perancangan system, implemantasi system, , pengujian dan analisis system dalam penelitain. Berikut ini adaaah langkah-langkah yang dilakukan untuk mengimplementasikan system “Implementasi Naive Byes Pada Proxy Server Untuk Klasifikasi Pengguna Internet - Naive Bayes Algorithm” :



Gambar 3.1 Diagram Alir Tahapan Penelitian

3.1 Study Literatur

Study literatur yang berhubungan dengan kasus atau penilitan yang dibahas, antara lain :

1. Teori tentang Karakteristik Pengguna
2. Teori tentang TFIDF
3. Teori Naïve Bayes Algorithm

Semua refrensi tersebut dapat dicari melalui buku, jurnal, artikel laporan penelitian, serta situs situs resmi di internet. Hasil yang diinginkan dari studi literatur tersebut adalah terkumpulnya semua koleksi refrensi yang relevan dengan rumusan masalah.

Tujuan dari studi literatur tersebut adalah untuk memperkuat masalah serta sebagai landasan teori untuk dapat melakukan studi dan juga menjadi dasar dalam melakukan sebuah perancangan dan Implementasi Naive Byes Pada Proxy Server Untuk Klasifikasi Pengguna Internet - *Naive Bayes Algorithm*.

3.2 Analisis Kebutuhan

Dalam tujuan kebutuhan dalam Implementasi Naive Byes Pada Proxy Server Untuk Klasifikasi Pengguna Internet - *Naive Bayes Algorithm* membutuhkan perangkat keras dan perangkat lunak untuk dapat mengimplementasikannya, antara lain :

1. Kebutuhan perangkat keras

Perangkat keras yang dibutuhkan dalam merancang serta mengimplementasikan topologi jaringan yang digunakan adalah :

a) Akses Point (AP)

Berfungsi dalam menghubungkan antara komputer dengan jaringan internet, perangkat yang digunakan adalah Tp-link

b) Switch

Berfungsi dalam penjemputan antara jaringan satu dengan jaringan lainnya, perangkat yang digunakan adalah Tp-link

c) Router board

Berfungsi dalam mengatur aktivitas pengguna internet di dalam suatu jaringan. Perangkat yang digunakan adalah Router Mikrotik memiliki 5 buah jumlah port , Ram 512Mb

d) Pc atau Laptop

Berfungsi sebagai server dalam mengambil data pada aktivitas pengguna internet, perangkat yang digunakan adalah Toshiba yang memiliki Ram 4Gb , HDD 100Gb, Intel, Processor Inter Centrino Duo @ 1.80GHz dan sebuah layar monitor 14" untuk mengumpulkan hasil keluaran (Output).

2. Kebutuhan perangkat lunak

Pada penelitian ini, dibutuhkan beberapa perangkat lunak yang sangat berperan dalam pengimplementasian dari analisis kriteria pengguna internet ini. Perangkat lunak yang di butuhkan adalah Sistem operasi Mikrotik yang sudah terinstal langsung dalam perangkat routerboard mikrotik , Linux ubuntu sebagai Squid Server sedangkan Windows 64bit sebagai pengeksekusi program Implementasi Naive Byes Pada Proxy Server Untuk Klasifikasi Pengguna Internet. JetBrains PhpStorm 7.1.3 sebagai *editor* dan *compiler* dari bahasa pemrograman HTML atau website

3.3 Pengumpulan Data

Dalam implementasi kriteria pengguna, tentu saja diperlukan sekumpulan data penting berupa informasi yang dapat digunakan sebagai data latih atau tolak ukur dalam penentuan pengkategorian dalam analisis menggunakan algoritma Naive Bayes Classifier (NBC):

1. Pengambilan Dataset

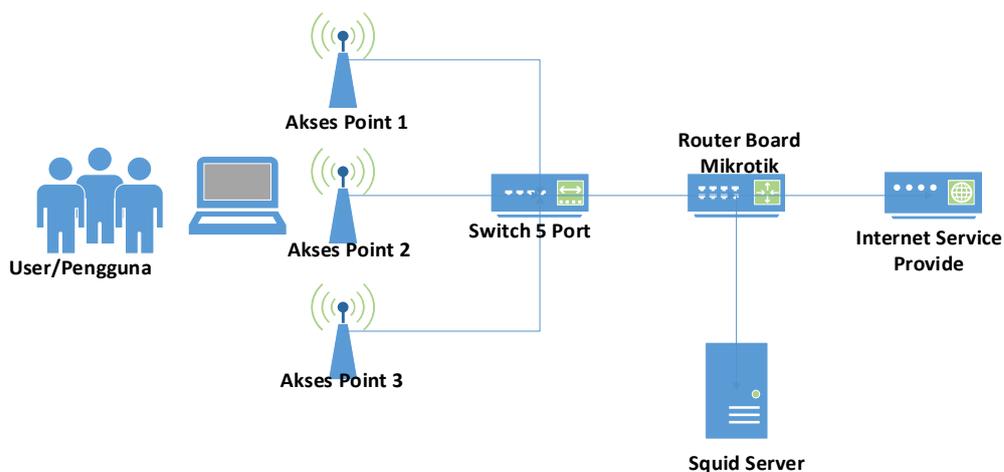
Data yang berfungsi sebagai atribut untuk data pembandingan pada pedoman Kamus Besar Bahasa Indonesia (KBBI) dengan alamat <http://kbbi.co.id/daftar-kata> dan bahasa inggris versi [oxfordlearnersdictionaries](http://www.oxfordlearnersdictionaries.com/wordlist/english/oxford3000/) dengan alamat <http://www.oxfordlearnersdictionaries.com/wordlist/english/oxford3000/>.

Jumlah total dataset dari Kamus Besar Bahasa Indonesia (KBBI) dan Bahasa Inggris berjumlah 35,315 kalimat.

2. Uniform Resource Locator (URL) Website
Uniform Resource Locator diambil di dalam Riwayat dari SARG squid yang sudah terbentuk pada BAB 4 **Gambar5.19 Tampilan SARG** .
3. Jenis Metadata Tag Website
Metadata tag diambil melalui pemeriksaan dari setiap Uniform Resource Locator yang terekamonline squid di dalam SARG.

3.4 Perancangan Sistem

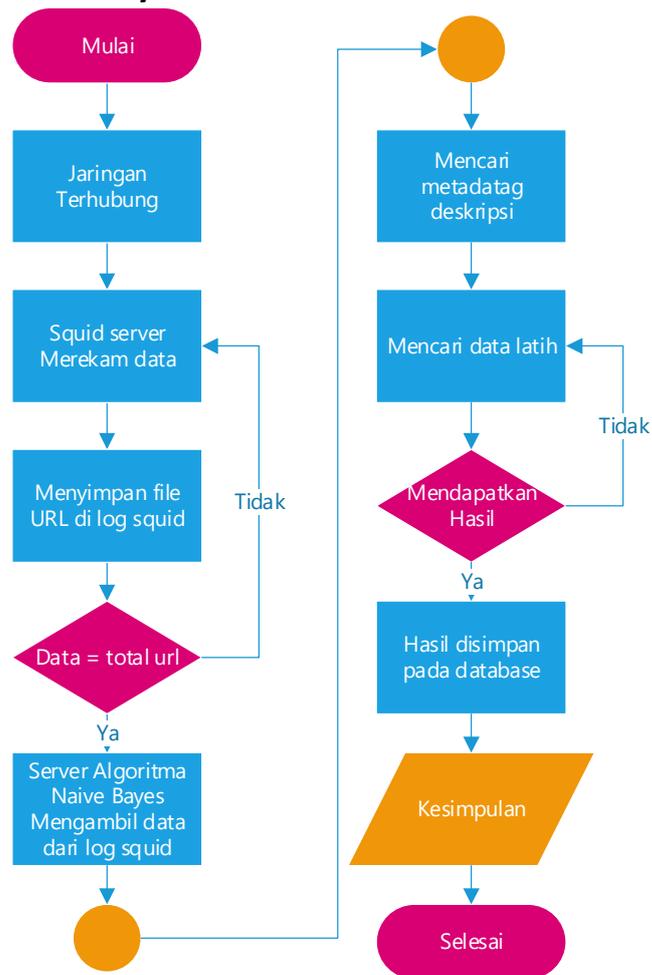
Perancangan dilakukan sesudah didapatkan seluruh data dari analisis kebutuhan yang sudah terkumpul. Dalam pembuatan sistem ini fokus dari perancangannya adalah antara router board, web proxy dan squid server. Gambar merupakan perencanaan topologi yang akan di buat:



Gambar 3.2 Topologi Jaringan

Dari perancangan ini akan dijelaskan lebih lanjut tugas serta spesifikasinya masing masing, yang kedepannya akan mempermudah konfigurasi masing masing perangkat pada tahap implementasi.

3.5 Algoritma Proses System



Gambar 3.3 Flowchart System

3.6 Proses Klasifikasi

Dalam proses klasifikasi data yang didapat adalah berupa nilai data yang di diambil dari aktifitas pengguna internet, sehingga data akan di proses datannya sesuai dengan alogritma *naive bayes*. Selain menggunakan cara ini, dapat juga menggunakan rumus *DateValue* pada *Excel* dimana konversi ini dapat menghitung dengan bentuk manual untuk setiap datannya. Saat menggunakan rumus pada *Excel* ini, akan dilakukan pengumpulan data terlebih dahulu untuk menjadi lebih terstruktur.

3.7 Implementasi

Program pada Implementasi Naive Byes Pada Proxy Server Untuk Klasifikasi Pengguna Internet yang akan dibuat ini memiliki cara kerja sebagai berikut :

1. Pengguna melakukan aktivitas normal pada browsing internet pada jaringan yang sudah di terapkan SARG (Squid Analysis Report Generator).
2. Program akan menyimpan detail url pada SARG (Squid Analysis Report Generator) yang nanti di simpan ke database squid.
3. Program akan mencari data metadata tag pada masing masing url yang berada pada database.
4. Program dapat melakukan preprocessing pada kalimat kalimat yang didapat.

5. Program nantinya dapat memproses text ke dalam pencarian bentuk Regular Expression(Regex).
6. Program dapat menghitung jumlah algoritma *naive bayes* dan menampilkan hasil perhitungan yang telah dilakukan.

3.8 Pengujian dan Analisis Sistem

Pengujian yang dilakukan adalah dengan menggunakan *Confusion Matrix*. Prosedur pengujian dengan melihat tingkat hasil Posterior pada algoritma *naive bayes* dan menyesuaikan dengan hasil manual yang didapat. Tahap ini dilakukan untuk melakukan evaluasi apakah hasil yang didapat sesuai dan akurat sesuai dengan kenyataan yang ada atau tidak.