

## **BAB I**

### **PENDAHULUAN**

#### **1.1 Latar Belakang**

Demam berdarah (DBD) menjadi salah satu penyakit yang sering terjadi di wilayah Indonesia dengan jumlah penderita yang sangat banyak dan luas persebarannya di tiap wilayah di Indonesia. Terbukti pada tahun 2017 kementerian kesehatan mendapatkan jumlah kasus demam berdarah (DBD) yang terjadi di seluruh Indonesia mencapai angka 68.407 kasus dengan jumlah kematian sebesar 493 [1]. Tingginya kasus demam berdarah tersebut disebabkan masih kurangnya kesadaran masyarakat terkait perilaku hidup bersih untuk pengendalian agar tidak terjangkit oleh penyakit demam berdarah.

Twitter adalah salah satu media sosial yang masih digunakan pada saat ini. Twitter memiliki fitur yaitu pengguna mampu mengirim dan juga membalas sebuah informasi atau pesan berbasis teks dengan batasan karakter sebanyak 140 karakter. Di Indonesia sendiri pengguna aktif dari media sosial Twitter ini sangatlah banyak hingga mencapai angka sebesar 72.800.000 pengguna [2]. Besarnya angka pengguna twitter di Indonesia memungkinkan untuk menggunakan Twitter sebagai salah satu media sosial yang efektif untuk pengguna mencari informasi khususnya informasi terkait penyakit demam berdarah (DBD) yang terjadi di Indonesia. Dengan demikian dari Twitter dapat dilakukan pengumpulan data terkait demam berdarah. Dari data yang telah terkumpul dapat dilakukan pengklasifikasian *tweet* yang menyatakan kasus demam berdarah (DBD) sekaligus memberikan informasi lokasi di Indonesia yang terjangkit demam berdarah (DBD).

Klasifikasi merupakan contoh masalah yang dapat diselesaikan dengan *text mining* [3]. Sistem klasifikasi dibuat dari data latih yang sudah ditentukan kelasnya yang digunakan untuk menentukan kelas dari data uji. Algoritma yang bisa digunakan untuk melakukan proses klasifikasi adalah Naïve Bayes, K-Nearest Neighbor, Support Vector Machine, C4.5, ID3, Artificial Neural Network [4]. Naïve Bayes merupakan metode yang dikemukakan oleh Thomas Bayes yang dapat digunakan

untuk melakukan klasifikasi dengan menerapkan perhitungan nilai probabilitas. Naïve Bayes merupakan algoritma yang mudah dalam implementasi, efisien dalam melakukan klasifikasi, dan tingkat prediksi yang cukup baik [4]. Hasil yang didapatkan ketika melakukan klasifikasi wabah penyakit demam berdarah dan chikungunya di New Delhi mendapatkan nilai mencapai 92% [5]. Proses klasifikasi pada Twitter memiliki kelemahan dikarenakan ada batasan dalam penulisan karakter sehingga pengguna sering menggunakan singkatan kata yang dapat mengakibatkan kurang tepatnya hasil klasifikasi sehingga perlu ditambahkan metode *n-gram* untuk menghindari ketidaktepatan pada proses klasifikasi [6]. Untuk mendapatkan lokasi terjadinya kasus demam berdarah dari *tweet* akan digunakan metode Named Entity Recognition (NER). NER merupakan metode yang digunakan untuk melakukan ekstraksi informasi yang belum terstruktur menjadi sebuah informasi yang terstruktur [7].

Dari penjelasan yang telah dijabarkan maka penulis memutuskan akan melakukan penelitian terkait membuat model yang kemudian diterapkan pada sistem untuk melakukan klasifikasi dan deteksi lokasi kasus demam berdarah di indonesia berdasarkan *tweet* dengan menggunakan metode Naïve Bayes dan Named Entity Recognition. Diharapkan hasil penelitian ini dapat memberikan informasi terkait daerah yang terjangkit penyakit demam berdarah di Indonesia sehingga dapat meningkatkan kesadaran masyarakat untuk melakukan pencegahan penyakit demam berdarah.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang diatas, maka rumusan masalah pada penelitian ini adalah sebagai berikut :

- a. Bagaimana implementasi sistem klasifikasi kasus demam berdarah di Indonesia dengan menggunakan data *tweet* sebagai sumber data klasifikasi?
- b. Bagaimana implementasi sistem deteksi lokasi demam berdarah di Indonesia dengan menggunakan hasil klasifikasi kasus demam berdarah?

### **1.3 Tujuan**

Tujuan pada penelitian ini adalah sebagai berikut :

- a. Implementasi model klasifikasi kasus demam berdarah di Indonesia dengan data yang diambil dari Twitter dengan menggunakan metode Naive Bayes.
- b. Implementasi model deteksi lokasi kasus demam berdarah di Indonesia berdasarkan hasil klasifikasi kasus demam berdarah pada data Twitter menggunakan Named Entity Recognition.

### **1.4 Batasan Masalah**

Berikut batasan masalah dalam penelitian adalah sebagai berikut :

- a. Dataset yang digunakan berasal dari Twitter dan data *tweet* yang diambil merupakan *tweet* yang menggunakan bahasa Indonesia.
- b. Pencarian *tweet* berdasarkan kata kunci “demam berdarah”, “dbd”, dan “dengue”.
- c. Deteksi lokasi kasus demam berdarah hanya mencakup wilayah Indonesia.

### **1.5 Sistematika Penulisan**

Untuk memberikan gambaran dari laporan tugas akhir, terdapat penjelasan bagian dari laporan yaitu Pendahuluan, Tinjauan Pustaka, Metodologi Penelitian, Hasil dan Pembahasan, serta Kesimpulan dan Saran. Kelima bagian laporan tugas akhir adalah sebagai berikut:

## **BAB I PENDAHULUAN**

Bab ini terdiri dari latar belakang dari penelitian, rumusan masalah yang ada pada penelitian, tujuan penelitian, batasan masalah dari penelitian, dan sistematika dari penulisan laporan.

## **BAB II LANDASAN TEORI**

Bab ini terdiri dari studi literatur yang membahas tentang referensi yang berhubungan dengan penelitian yang dikerjakan dan juga landasan teori yang

membahas tentang teori-teori terkait dengan pembahasan penelitian yang akan dikerjakan.

### **BAB III METODOLOGI PENELITIAN**

Bab ini terdiri dari alur penelitian, gambaran umum dari keseluruhan sistem yang akan dibuat, dan perancangan sistem untuk mempermudah tahap implementasi.

### **BAB IV HASIL DAN PEMBAHASAN**

Bab ini terdiri dari implementasi yang dilakukan pada penelitian dan hasil yang didapatkan setelah melakukan proses implementasi.

### **BAB V KESIMPULAN DAN SARAN**

Bab ini terdiri dari kesimpulan yang didapat dari hasil perancangan dan implementasi pada penelitian serta saran untuk pengembangan selanjutnya dari penelitian yang sudah dikerjakan.