

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Perkembangan teknologi informasi dan komunikasi *digital* saat ini memudahkan masyarakat untuk saling bertukar informasi satu sama lain. Hal tersebut mengubah kebiasaan masyarakat dalam berbagai bidang. Tidak terkecuali perkembangan tersebut juga mempengaruhi perkembangan media bertukar informasi. Media bertukar informasi yang saat ini paling banyak digunakan yaitu media sosial. *Twitter* merupakan salah satu jejaring sosial yang banyak digunakan saat ini. Di Indonesia tercatat hingga tahun 2018 pengguna *twitter* mencapai 130 juta pengguna aktif [1]. *Twitter* merupakan jejaring sosial terbuka yang memungkinkan pengguna untuk melacak pengiriman (disebut tweets) dari pengguna lain tanpa mendapatkan persetujuan.

Kondisi tersebut memberikan kesempatan kepada perkembangan penelitian untuk mempelajari tingkah laku masyarakat. Salah satunya peneliti mulai untuk menganalisis kebiasaan penderita depresi ketika menggunakan sosial media yang menyediakan kesempatan kepada penggunanya untuk dapat mengekspresikan perasaan secara anonim dan rahasia. Melalui data teks *tweet* pengguna, data yang mengandung kalimat depresif dapat diidentifikasi dan diklasifikasikan dengan menggunakan banyak metode.

Terdapat beberapa penelitian klasifikasi teks menggunakan metode *stastical based* telah banyak dilakukan seperti pendeteksian depresi pada media

sosial yang dilakukan oleh Al-Asad, Pranto, Afreen, dan Islam dengan menggunakan *naïve bayes* untuk pengklasifikasian data tweet pengguna, dengan akurasi 74%[2]. Penelitian lainnya yaitu pengklasifikasian komentar cyberbullying dengan SVM [3], pengklasifikasian kepribadian berdasarkan teks twitter menggunakan *naïve bayes*, KNN dan SVM[4], pengukuran tingkat depresi pada sosial media menggunakan *Rule based dan Statistical based*[5], pendeteksian depresi menggunakan multinasional *naïve bayes* dan support vector regression[6], pengklasifikasian teks lirik menggunakan *naïve bayes*[7], dan pengklasifikasian berita menggunakan multinomial dan bernouli *naïve bayes*[8]. Adapun penelitian pengklasifikasian lain dengan metode *rule based* salah satunya yaitu penelitian yang dilakukan oleh Nurjannah, pada penelitian sistem klasifikasi teks gejala depresi dengan menggunakan *rulebased* algoritma *Synescatch* dengan akurasi sebesar 86,91%[9], adapun penelitian lainnya oleh Atmaja dan Purwantianti pada penelitiannya mengenai perbandingan *rule based* dan *statistical based method* pada pengklasifikasian emosi dengan menggunakan data tweet berbahasa Indonesia. Pada penelitian tersebut Atmaja dan Purwantianti menggunakan algoritma *synescatch* untuk metode *rule based*, dan didapatkan akurasi sebesar 62.070 % [10]

Berdasarkan penelitian sebelumnya pada pengklasifikasian teks menggunakan metode *statistical based* dan *rule based*, maka penulis tertarik membangun sistem pengklasifikasian kalimat depresif untuk membandingkan akurasi dari algoritma kedua metode tersebut. Secara algoritma, metode *rule based* berarti aturan klasifikasi dilakukan secara manual berdasarkan aturan yang dibuat, sedangkan metode *statistical based* berarti aturan klasifikasi ditentukan secara

otomatis oleh algoritma pembelajaran mesin berdasarkan data yang ada. Maka penulis melakukan penelitian mengenai ***Perbandingan Algoritma Statistical Based dan Rule Based pada Analisis Teks Kalimat Depresif di Twitter.***

1.2. Rumusan Masalah

Rumusan masalah yang menjadi dasar penelitian perbandingan algoritma statistical based dan rule based pada kalimat depresif di *twitter* ini berdasarkan latar belakang yang telah kita jabarkan adalah sebagai berikut.

1. Bagaimana menerapkan metode *rule based* dan *statistical based*?
2. Bagaimana perbandingan kinerja metode *rule based* dan *statistical based*?

1.3. Tujuan

Adapun tujuan penelitian ini yang didasarkan pada uraian latar belakang di atas adalah sebagai berikut.

1. Menerapkan metode *rule based* dan *statistical based*.
2. Mengetahui perbandingan kinerja metode *rule based* dan *statistical based*

1.4. Batasan Masalah

Penelitian memiliki batasan masalah yang dapat diuraikan sebagai berikut.

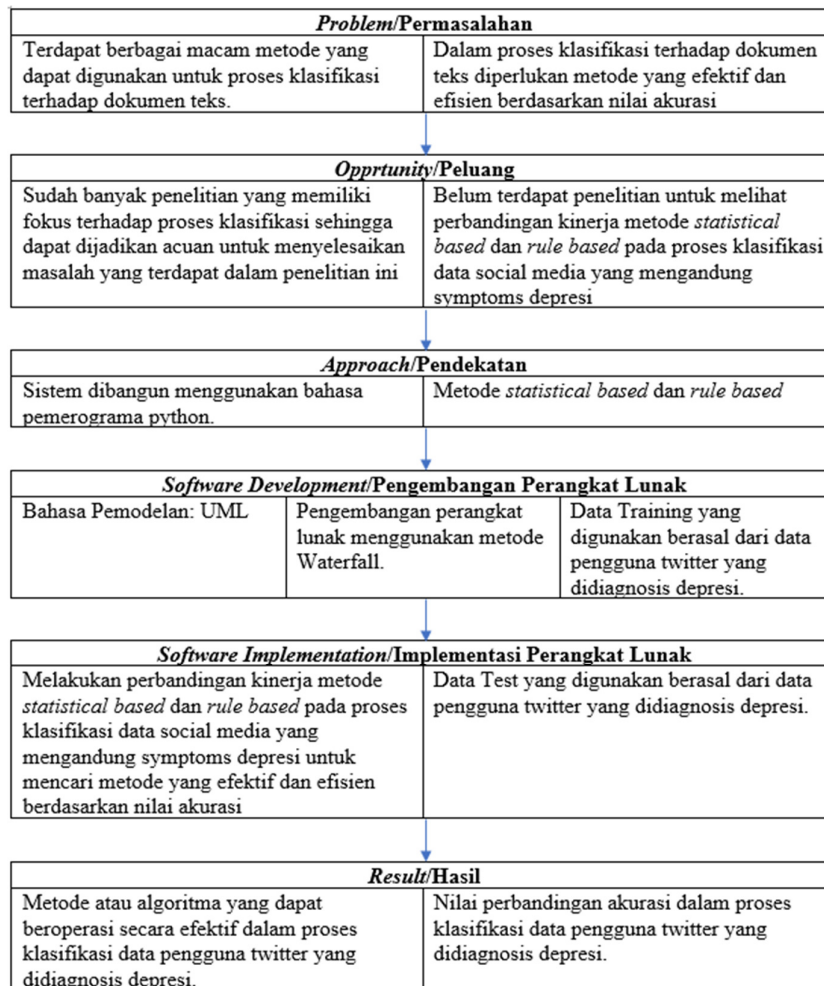
1. Pengembangan sistem menggunakan bahasa pemrograman *Python*.
2. Data yang dapat diporses oleh sistem klasifikasi kalimat depresif ini berupa teks.

3. Data yang diambil adalah data penderita depresi yang pernah menuliskan “Aku/gue/saya didiagnosis depresi...”
4. Data yang diambil dibatasi sebanyak 3000 data.
5. Data diambil menggunakan metode crawling dengan *Twitter* API
6. Parameter yang menjadi fokus penelitian ialah nilai akurasi yang dihasilkan dari proses klarifikasi data akun twitter yang didiagnosis depresi.



1.5. Kerangka Pemikiran

Adapun kerangka pemikiran dalam penelitian yang akan kami lakukan dapat digambarkan pada Gambar 1.1 di bawah ini.



Gambar 1. 1 Kerangka Pemikiran

Gambar di atas menunjukkan kerangka pemikiran dari penelitian yang akan penulis lakukan, terdiri dari aspek permasalahan yaitu untuk menemukan metode yang paling tepat untuk klasifikasi teks, dalam penelitian sebelumnya terdapat penelitian klasifikasi teks dengan metode *rule based* dan *statistical based* sehingga menciptakan peluang bagi penulis untuk membuat penelitian mengenai

perbandingan kedua metode tersebut. Adapun pendekatannya melalui bahasa pemrograman python, dengan proses pengembangan menggunakan *Unified Modeling Language (UML)* untuk perancangan dan pemodelannya, sedangkan metode pengembangan menggunakan metode *waterfall* dengan data yang digunakan sebanyak 3000 *dataset* dan 1500 *data training* serta 750 *data testing* dari tweet pengguna yang mengalami depresi. Sehingga, kita mendapat hasil akhir berupa perbandingan metode *rule based* dan *statistical based* pada kalimat depresif di twitter.

1.6. Metodologi Penelitian

Adapun metodologi penelitian yang penulis terapkan untuk mendapatkan data serta informasi dalam menguraikan masalah antara lain sebagai berikut.

a. Studi Literatur

Pada proses ini penulis mengumpulkan referensi pustaka yang bersumber dari buku, artikel, jurnal, makalah, maupun situs internet.

b. Analisis

Pada tahap analisis penulis menganalisis berbagai hal untuk kepentingan penelitian antara lain sebagai berikut.

- 1) Menganalisis tahapan-tahapan proses klasifikasi teks.
- 2) Menganalisis cara kerja dari metode *statistical based* dan *rule based* dalam klasifikasi kalimat depresif.

c. Perancangan

Adapun perancangan yang dilakukan untuk penelitian ini adalah perancangan arsitektur, data dan antarmuka.

d. Pengkodean

Adapun pada tahap ini merupakan tahapan implementasi pengkodean program sistem klasifikasi kalimat depresif dengan menggunakan bahasa pemrograman yang sudah ditentukan.

e. Pengujian

Tahap pengujian akan dilakukan untuk memastikan bahwa sistem berjalan sesuai dengan spesifikasi rancangan.

f. Penyusunan Laporan

Penyusunan laporan dilakukan untuk mendokumentasikan semua tahapan analisis, implementasi serta hasil yang didapat dari penelitian ini.

1.7. Sistematika Penulisan

Dalam penulisan tugas akhir ini disusun berdasarkan sistematika penulisan yang terdiri dari lima bab. Berikut dengan penjelasannya:

BAB I PENDAHULUAN

Pada bab ini dijelaskan latar belakang, rumusan masalah, tujuan dan manfaat penelitian, batasan masalah, metode penelitian dan sistematika penelitian.

BAB II LANDASAN TEORI

Pada bab ini penulis akan menguraikan landasan teori berupa tinjauan pustaka serta landasan teori dari penelitian yang penulis lakukan.

BAB III ANALISA DAN PERANCANGAN

Pada bab ini penulis akan membahas analisis serta perancangan sistem yaitu analisis sistem, data, metode, arsitektur sistem, pemodelan sistem, serta rancangan database, antar muka dan pseudocode algoritma.

BAB IV IMPLEMENTASI DAN PENGUJIAN

Pada bab ini berisi dokumen implementasi sistem diawali dengan pembahasan mengenai lingkungan implementasi, lalu implementasi sistem hingga pengujian sistem.

BAB V PENUTUP

Pada bab ini dijelaskan simpulan dari penelitian yang dilakukan dan berisi saran sebagai masukan untuk penelitian dan pengembangan selanjutnya.

