

BAB I PENDAHULUAN

1.1. Latar Belakang

Sosial media merupakan suatu hal yang sudah tidak bisa dipisahkan lagi dengan masyarakat, pada suatu survei menyatakan bahwa 97,4% masyarakat Indonesia menggunakan media sosial (Handikasari et al., 2018). Dengan kebebasan pengguna Twitter dalam menuliskan tweet-tweet nya, untuk sebagian pengguna hal-hal yang cukup mendalam dalam kehidupan pribadinya pun sering kali dibagikan agar dapat bisa dilihat oleh khalayak banyak. Salah satu contoh hal pribadi yang sering dibagikan secara tidak sadar di bagikan adalah mengeneai kesehatan mental si pengguna. Dari penggunaan kata seorang pengguna Twitter dapat menunjukkan bahwa ia sedang mengalami salah satu dari kesehatan mental yaitu Depresi.

Pada awal 2019 dilansir oleh Suara.com, seorang pria berumur 21 tahun ditemukan bunuh diri setelah menuliskan kegelisahanya di akun twitter miliknya. Dalam tweet nya pria ini menuliskan kata-kata berbau depresi selama 12 hari terakhir, hingga- akhir nya ia ditemukan tewas menenggak racun *arsenic* dikamar kosnya. Sebelum ia bunuh diri ia menuliskan pada tweet nya “gue gamau terus menerus jadi benalu di lingkungan gua, so kayanya lebih baik gua enyah dari muka bumi ini. bye!”. Dari tweet ini menunjukkan bahwa pria ini akan segera mengakhiri hidupnya dengan bunuh diri (Oktarini D.S., 2019).

Layanan kesehatan jiwa yang salah satunya adalah depresi ini masih sangat jarang di Indonesia, bahkan Indonesia menempati peringkat terendah dalam penyediaan layanan kesehatan jiwa ini (Wulandari et al., 2015). Masalah kesehatan mental depresi ini tentu bukanlah hal yang ringan mengingat akibat dari depresi yang berlebihan ini dapat membuat seseorang untuk bunuh diri (Marcus et al., 2012). Pada penelitian yang menjelaskan tentang hubungan antara media sosial dan depresi pada mahasiswa menunjukkan semakin tinggi intensitas penggunaan media sosial, semakin tinggi pula derajat depresi yang dialami mahasiswa. (Handikasari et al., 2018). Seiring dengan maraknya penggunaan social media dan terjadinya kasus penggunaan twitter yang menuliskan kegelisahannya di akun twitter nya seperti dilansi oleh (Oktarini D.S, 2019).

Pada penelitian yang dilakukan oleh (M. I. Maulana & Soebroto, 2019) mengenai Klasifikasi Tingkat Stres Berdasarkan Tweet pada Akun Twitter. Klasifikasi teks yang digunakan pada penelitian ini adalah metode *Improved k-Nearest Neighbor* yang merupakan improvisasi dari metode *k-Nearest Neighbor* yang mempunyai kelemahan dalam penggunaan nilai k terhadap semua kelas. Selain itu, untuk meningkatkan akurasi dari sistem serta menghilangkan fitur yang kurang relevan maka digunakan seleksi fitur *chi-square* yang dapat menghilangkan fitur yang kurang relevan tanpa menurunkan akurasi dari sistem. Pada penelitian ini pun memiliki topik yang hampir sama yaitu mengenai *mental health*. Kemudian pada penelitian ini menyarankan

untuk penelitian selanjutnya agar dapat melakukan metode seleksi fitur yang berbeda.

Pada penelitian mengenai analisis sentiment dan klasifikasi kategori terhadap tokoh publik pada twitter yang dilakukan oleh (Fathan Hidayatullah & Sn, 2014) komparasi Algoritma Klasifikasi antara *Support Vector Machine* dengan *Naïve bayes* menghasilkan hasil bahwa Algoritma *Support Vector Machine* menghasilkan akurasi performansi yang lebih baik dibandingkan dengan Algoritma *Naïve bayes* dalam hal Klasifikasi Teks. Kemudian pada penelitian lainnya yang dilakukan oleh (Asiyah & Fithriasari, 2016) mengenai komparasi Algoritma Klasifikasi teks antara *Support Vector Machine* dengan *K-Nearest Neighbor* untuk klasifikasi berita online menunjukkan bahwa performansi *Support Vector Machine* jauh lebih baik daripada *K-Nearest Neighbor* dalam hal klasifikasi teks.

Pada Penelitian yang dilakan oleh (Simanjuntak, 2017) Metode SVM ini merupakan salah satu metode yang sering digunakan untuk klasifikasi pada teks dalam text mining seperti telah dijelaskan sebelumnya berdasarkan 2 penelitian mengenai komparasi algoritma untuk klasifikasi teks. SVM merupakan salah satu metode terbaik dalam komputasi yang efisien dan memiliki *performance predictive* yang baik pula.

Seleksi fitur adalah proses untuk membuang fitur-fitur yang tidak atau kurang relevan pada proses klasifikasi text (M. I. Maulana & Soebroto, 2019).

Pada penelitian mengenai implementasi seleksi fitur *Information Gain* pada algoritma klasifikasi *machine learning* untuk prediksi performa akademik siswa (Sari, 2016), penelitian ini menunjukkan bahwa dengan implementasi seleksi fitur *Information gain* dapat mempengaruhi hasil dari algoritma klasifikasi *Machine learning* (*J48*, *Random Forest*, *MLP*, *SVM (SMO)*, dan *Naïve Bayes*).

Berdasarkan seluruh uraian tersebut, masih terdapat peluang pengembangan penelitian tentang penggunaan seleksi fitur menggunakan *Information Gain* dan *Support Vector Machine* sebagai algoritma klasifikasi text. Maka dari itu pada penelitian ini akan membahas tentang “Seleksi fitur menggunakan *Information gain* dan metode *Support Vector Machine* untuk klasifikasi tingkat depresi pada data Twitter”.

1.2. Rumusan Masalah

Adapun rumusan masalah dari latar belakang di atas adalah :

- a. Bagaimana penerapan seleksi fitur *information gain* pada dataset tentang depresi pada pengguna twitter?
- b. Bagaimana penerapan metode *Support Vector Machine* dalam klasifikasi tingkat depresi pada dataset yang telah dilakukan seleksi fitur *Information Gain*?

1.3. Batasan Masalah

Beberapa hal yang menjadi batasan dan dasar dalam penelitian ini adalah :

- a. Metode yang digunakan dalam penelitian ini adalah Support Vector Machine dan seleksi fitur Information Gain untuk klasifikasi tingkat depresi pada tweet media social Twitter.
- b. Hanya melakukan klasifikasi dari metode *Support Vector Machine* dan seleksi fitur *Information Gain*.
- c. Data set yang digunakan hanya sebanyak 26.000 tweet yang di diambil pada rentang waktu dari tanggal 27 – 30 Juni 2021.
- d. Data set yang telah di dapat kemudian dilakukan *preprocessing* teks dan *labeling* agar dapat dilakukan proses pengklasifikasian.

1.4. Tujuan Penelitian

Adapun tujuan penelitian adalah:

- a. Mengetahui performa berdasarkan nilai akurasi dari penggunaan Seleksi Fitur *Information Gain* pada data set tentang depresi pada pengguna

Twitter dengan menggunakan algoritma klasifikasi *Support Vector Machine*.

1.5. Manfaat Penelitian

Adapun manfaat dari penelitian ini adalah :

- a. Dapat membantu untuk identifikasi dini dari pengguna twitter yang sedang mengalami depresi.
- b. Dapat mengklasifikasikan tingkat depresi dari pengguna twitter
- c. Mengetahui kinerja algoritma SVM dan seleksi fitur *Information Gain* untuk proses klasifikasi dalam penelitian ini.

1.6. Metodologi Penelitian

Pada penelitian ini diawali dengan pengambilan data tweet dari pengguna Twitter dengan API yang telah disediakan oleh Twitter secara gratis untuk tujuan penelitian dan pembelajaran. Pada proses pengambilan data ini akan menggunakan aplikasi Orange data mining untuk memudahkan proses pengambilan data. Selanjutnya untuk melakukan klasifikasi tingkat depresi pada tweet dari pengguna twitter akan menggunakan algoritma klasifikasi SVM.

Tahapan penelitian yang digunakan dalam menyelesaikan penelitian ini diantaranya :

- a. Teknik Pengumpulan Data
 1. Studi Pustaka

Studi pustaka dilakukan dengan mempelajari tentang teori–teori yang menjadi referensi dan pendukung dalam pembuatan dan penulisan hasil penelitian mengenai melakukan klasifikasi tingkat depresi berdasarkan tweet pada twitter dengan menggunakan algoritma Support Vector Machine (SVM) dan seleksi fitur *Information Gain*.

2. Data Scraping

Data scraping adalah proses ekstraksi data dari sebuah web. Pada penelitian ini data didapat langsung dari web Twitter. Twitter telah menyediakan API nya untuk mempermudah pengguna untuk mengakses informasi dengan terlebih dahulu mendaftar sebagai developer pada Twitter.

b. Metode Pengolahan Data

1. Text Mining

Text mining adalah teknik atau proses yang digunakan untuk eksplorasi, analisis, dan klasifikasi pada sekumpulan data tekstual yang besar yang digunakan untuk menemukan pola (Ratniasih et al., 2017). Pada penelitian kali ini text mining digunakan untuk analisis dan klasifikasi pada data tweet yang mengandung kalimat depresi

2. Text Preprocessing

Text preprocessing adalah suatu proses pengubahan bentuk data yang belum terstruktur menjadi data yang lebih terstruktur sesuai dengan kebutuhan, untuk proses *mining* lebih lanjut.

3. Seleksi fitur *Information Gain*

Seleksi fitur adalah teknik untuk memilih fitur penting dan relevan terhadap data dan mengurangi fitur yang tidak relevan atau seleksi fitur juga dapat diartikan sebagai teknik reduksi dimensi yang digunakan untuk memperkecil matriks data dengan memperhatikan informasi kata penting yang perlu diproses. Seleksi fitur bertujuan untuk memilih fitur terbaik dari suatu kumpulan data fitur (Maulida et al., 2016). *Information gain* merupakan salah satu metode seleksi fitur yang banyak dipakai oleh peneliti untuk menentukan batas dari kepentingan sebuah fitur (M. R. Maulana & Karomi, 2015). *Information Gain* merupakan teknik seleksi fitur yang memakai metode scoring untuk nominal ataupun pembobotan atribut kontinu yang didiskretkan menggunakan maksimal entropy.

4. Algoritma Klasifikasi *Support Vector Machine (SVM)*

SVM merupakan salah satu metode yang biasanya digunakan untuk klasifikasi. Dalam pemodelan klasifikasi, SVM memiliki konsep yang lebih matnag dan lebih jelas secara matematis.

c. Penarikan Kesimpulan

1.7. Sistematika Penulisan

Sistematika penulisan Tugas Akhir ini disusun untuk memberikan gambaran umum mengenai proses yang akan dijalankan. Adapun sistematika yang akan diuraikan secara garis besar, yaitu :

BAB I PENDAHULUAN

Berisikan tentang latar belakang penulisan Tugas akhir, Rumusan masalah, Ruang Lingkup penulisan, Tujuan dan Manfaat yang di dapat didapat dari hasil penelitian, serta metodologi, dan sistematika penulisannya.

BAB II TINJAUAN PUSTAKA

Dalam Bab ini berisi tentang konsep dasar dan teori-teori dasar mengenai *text mining*, Seleksi fitur, algoritma klasifikasi dan depresi yang akan dijadikan sebagai acuan dan landasan dalam penulisan Tugas Akhir ini.

BAB III METODOLOGI

Pada bab ini akan membahas mengenai tahapan-tahapan dalam melakukan seleksi fitur *Information gain* dan klasifikasi tingkat depresi berdasarkan tweet pada twitter dengan menggunakan algoritma Support Vector Machine (SVM).

BAB IV HASIL DAN PEMBAHASAN

Dalam bab ini berisikan tentang analisis dari hasil penelitian mengenai identifikasi tweet yang mengandung kata kata depresi, seleksi fitur *Information gain*, klasifikasi dari tingkat depresi yang telah di identifikasi menggunakan algoritma *Support Vector Machine*.

BAB V SIMPULAN DAN SARAN

Pada bab ini berisikan simpulan dari penelitian yang telah dilakukan dan saran yang dibutuhkan terhadap hasil akhir untuk penelitian selanjutnya.