

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pada era digital saat ini terjadi perkembangan teknologi dengan sangat cepat, menjadikan manusia mengalami banyak perubahan dalam menjalani kehidupan. Utamanya dalam hal komunikasi dan penyampaian pendapat. Perkembangan teknologi informasi khususnya adanya media sosial telah mengubah cara manusia dalam hal berkomunikasi antara satu sama lain. Masyarakat Indonesia banyak menggunakan media sosial untuk mengungkapkan pengalaman, pendapat, dan hal-hal lain yang menjadi perhatian mereka. Akhir-akhir ini, seiring meningkatnya peran penggunaan media sosial di masyarakat untuk berbagi informasi akan mempengaruhi perkembangan opini publik,

Salah satu platform media sosial paling populer di Indonesia adalah Twitter. Twitter dibuat oleh Jack Dorsey pada bulan Maret tahun 2006 yang merupakan *mikroblog online* yang memiliki fitur untuk pengguna dapat mengirim dan membaca pesan yang terdiri dari 140 karakter, yang dikenal sebagai kicauan (*tweet*) [1]. Menurut laporan Statista, hingga Januari 2022, Twitter mempunyai total pengguna aktif harian sekitar 206 juta pengguna yang tersebar di seluruh dunia, sementara Indonesia dengan total 18,45 juta pengguna menjadikan Indonesia berada di peringkat ke-5, yang persis satu tingkat lebih rendah dari Brazil 19,05 juta pengguna dan satu tingkat lebih tinggi dari 18,4 juta pengguna United Kingdom [2]. Basis utama Twitter sebagai platform *microblogging* yang berbasis teks menjadikannya sebagai media yang ideal untuk melakukan analisis sentimen tekstual, berbeda dengan situs jejaring sosial lainnya, seperti Instagram atau Youtube yang titik tekannya pada tampilan visual [3].

Twitter umumnya digunakan oleh pengguna untuk berbagai tujuan seperti sarana menyampaikan pendapat baik yang bersifat positif atau negatif, kampanye politik, alat pembelajaran, dan sebagai media komunikasi darurat [4]. Termasuk didalamnya opini masyarakat mengenai varian virus baru dari Covid-19 yaitu virus Omicron. Pada masa pandemi di Indonesia, berbagai usulan dan regulasi telah diperkenalkan sebagai strategi pencegahan dan penanganan penyebaran Covid-19, seperti anjuran protokol kesehatan dengan jaga jarak dan penggunaan masker, pembatasan perjalanan, pembatasan wilayah, dan vaksinasi. Sayangnya, Covid-19 terus menjadi masalah sejak munculnya varian baru yang menjadi perhatian, varian terakhir yang ditemukan adalah yang dikenal sebagai Omicron [5]. Varian ini juga dikenal dengan varian B.1.1.529 dan pertama kali muncul di Afrika Selatan pada tanggal 24 November 2021 [6]. Menteri Kesehatan Budi Gunadi Sadikin mengumumkan temuan kasus pertama

COVID-19 varian omicron di Indonesia pada tanggal 16 Desember 2021 [7]. Omicron menjadi virus varian baru yang penyebarannya lima kali lebih cepat dibandingkan dengan varian sebelumnya termasuk varian Delta [8]. Daya transmisi yang kuat tersebut menjadikan omicron ditetapkan sebagai *Variant of Concern*(VoC) [9]. Munculnya varian baru dari Covid-19 membuat masyarakat banyak menyampaikan opini yang lebih bebas melalui *tweet* di Twitter dengan mengungkapkan pendapat mereka mengenai Omicron. Meningkatnya penyajian ide tersebut dapat menjadi salah satu media untuk menganalisis kecenderungan informasi tentang suatu subjek menjadi positif atau negatif. Analisis kecenderungan ekspresi dari opini biasa disebut dengan analisis sentimen. Peneliti ingin mengetahui kecenderungan opini *tweets* mengenai kata kunci virus Omicron melalui analisis sentimen. Analisis sentimen merupakan proses otomatis untuk memahami, mengekstraksi dan memproses data tekstual untuk mendapatkan informasi pendapat yang terdapat dalam suatu kalimat opini [10].

Banyak metode *text mining* yang bisa diterapkan dalam analisis sentimen yaitu *Support Vector Machine*, *Lexicon Based*, *Naïve Bayes Classifier*, dan *K-Nearest Neighbor*. Dari beberapa metode *text mining*, metode yang diterapkan dalam penelitian ini yaitu *Support Vector Machine* dengan input data *tweet* dalam Bahasa Indonesia. SVM adalah metode yang algoritma klasifikasinya termasuk kelas *supervised learning*(*machine learning*) yang memprediksi kelas berdasarkan pola atau model dari hasil proses pelatihan [11]. *Supervised learning* adalah jenis kelas yang dalam penerapannya memerlukan tahapan pelatihan dan pengenalan pada suatu objek yang akan dianalisis serta adanya tahap pengujian [12]. SVM bertujuan mendapatkan *hyperplane* terbaik diantara dua buah kelas. SVM mempunyai kelebihan diantaranya menggunakan *Support Vector Machine* dalam penentuan jarak membuat proses komputasi menjadi cepat dan menunjukkan tingkat akurasi yang lebih baik dari pada pendekatan lain [1]. Seperti pada penelitian sebelumnya yang dilakukan oleh Lutfi dan Muhammad yang melakukan perbandingan klasifikasi review tekstual di *Google Play Store* menggunakan *Support Vector Machine* dan *Naive Bayes* untuk Analisis Sentimen, serta hasil pengujian menggunakan teknik *3-Folds Cross Validation*, klasifikasi ulasan teks bahasa Indonesia di *Google Play Store* memiliki skor yang lebih tinggi untuk klasifikasi menggunakan SVM dibandingkan dengan akurasi klasifikasi menggunakan *Naive Bayes*. Artinya, pengklasifikasi SVM menerima akurasi sebesar 81,46% sedangkan pengklasifikasi *Naive Bayes* sebesar 75,41% [13].

Dalam penelitian lain yang dilakukan oleh Fajar dan Iqbal metode SMV juga menunjukkan tingkat skor akurasi yang lebih tinggi dibandingkan dengan metode yang lain, metode *Support Vector Machine* (SVM), *Naive Bayes*, dan *K-Nearest Neighbor* digunakan untuk menganalisis sentimen reaksi masyarakat Indonesia terhadap pandemi Covid-19 di media sosial

Twitter, dan tiga algoritma yang paling baik dalam mengklasifikasikan data opini dibandingkan. Algoritma SVM dengan kernel linier memiliki akurasi yang lebih tinggi dengan akurasi rata-rata 90,01% dibandingkan dengan *Naive Bayes* dengan akurasi rata-rata 79,20% dengan jumlah *laplace* 1, dan KNN dengan kernel optimal memiliki akurasi rata-rata 62,10 % dengan jumlah K adalah 20 berdasarkan tingkat akurasi rata-rata dari evaluasi model *10-fold cross validation* [14].

Hanya saja SVM juga memiliki kekurangan yaitu sulit digunakan untuk mengolah sample dalam jumlah besar serta SVM secara teoritik hanya dikembangkan untuk klasifikasi dua kelas. Masalah lain dengan kategorisasi analisis sentimen SVM adalah banyaknya jumlah fitur yang terdapat dalam kumpulan data yang dapat menyebabkan nilai akurasi yang dihasilkan kurang maksimal. Untuk mengatasi kekurangan ini, jumlah fitur harus dikurangi untuk dapat meningkatkan ketepatan atau nilai akurasi dari metode SVM dengan seleksi fitur. Dalam penelitian ini menggunakan *Partical Swarm Optimization*(PSO) untuk seleksi fitur. Dikarenakan pada penelitian sebelumnya menunjukkan bahwa teknik PSO menghasilkan nilai akurasi yang lebih unggul untuk digunakan pada metode SVM dibandingkan metode yang lain. Yaitu pada penelitian yang dilakukan oleh Beni Rahmatullah, Pungkas Budiyo dan Suwanda Aditya Saputra yang menganalisis sentimen transportasi online di twitter yang menggunakan algoritma *Naive Bayes*, SVM dan KNN sebagai perbandingan dengan menggunakan *Particle Swarm Optimization* untuk seleksi fiturnya, dan hasil akurasi menunjukkan dengan seleksi fitur PSO menambah nilai akurasi serta berdasarkan hasil yang diperoleh algoritma SVM dengan PSO memiliki skor ketepatan paling tinggi dengan akurasi 98% [15].

Pada penelitian ini dilakukan pengukuran tingkat akurasi dari penggunaan metode *Support Vector Machine (SVM)* dalam menganalisis sentimen *tweet* mengenai virus Omicron di Indonesia pada media sosial Twitter menggunakan seleksi fitur *Particle Swarm Optimization*.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan sebelumnya, maka berikut dapat diperoleh perumusan masalah dari penelitian ini:

Bagaimana tingkat akurasi dari penerapan metode *Support Vector Machine (SVM)* dengan seleksi fitur *Particle Swarm Optimization* dalam mengklasifikasikan sentimen opini masyarakat melalui *tweet* di media sosial Twitter mengenai virus Omicron yang ada di Indonesia?

1.3 Tujuan dan Manfaat

1.3.1 Tujuan

Berdasarkan rumusan masalah yang telah diuraikan, dapat diketahui bahwa tujuan dari penelitian ini yaitu sebagai berikut :

Mengetahui tingkat akurasi dari penerapan metode *Support Vector Machine (SVM)* dengan seleksi fitur *Particle Swarm Optimization (PSO)* dalam mengklasifikasikan sentimen opini masyarakat melalui *tweet* di media sosial Twitter mengenai virus Omicron yang ada di Indonesia.

1.3.2 Manfaat

1. Manfaat Teoritis

Diharapkan adanya penelitian ini dapat membantu menambah ilmu dan pengetahuan mengenai analisis sentimen publik terhadap virus omicron di media sosial Twitter.

2. Manfaat Praktis

- Dapat mengetahui serta mengklasifikasikan ekspresi dari opini publik terhadap virus Omicron.
- Tingkat akurasi pendekatan *Support Vector Machine* dalam analisis sentimen dapat diketahui.
- Dapat mengetahui tingkat akurasi dari metode *Support Vector Machine* setelah penambahan seleksi fitur *Particle Swarm Optimization*.

1.4 Batasan Masalah

Batasan masalah diharapkan mampu menjadikan penelitian lebih terarah dari pada tersebar secara umum, dan sebagai penentu area diskusi tertentu. Berikut merupakan batasan masalah yang ada pada penelitian ini:.

1. Klasifikasi hanya menggunakan kelas positif dan kelas negatif.
2. Data yang digunakan diperoleh dari media sosial Twitter.
3. Kicauan (*tweet*) yang dianalisis hanya *tweet* berbahasa Indonesia.
4. Data yang digunakan hanya data dengan kata kunci Virus Omicron
5. Data *tweet* yang digunakan adalah data selama empat bulan semenjak munculnya virus Omicron yaitu dengan selang waktu dari 27 November 2021 sampai dengan 27 Maret 2022.
6. Data yang digunakan hanya sebanyak 1000 *tweet* dengan 500 *tweet* positif dan 500 *tweet* negatif.

7. Metode yang digunakan yaitu *Support Vector Machine (SVM)* dengan seleksi fitur *Particle Swarm Optimization (PSO)*.
8. Menggunakan bahasa pemrograman Python.
9. Pengambilan data *tweet* dilakukan dengan *Scraping* data dengan *Library python*.
10. Implementasi pada tahapan *filtering* dan *stemming* menggunakan *Library Sastrawi*.

1.5 Sistematika Penulisan

Terdapat beberapa bab yang tersusun dalam sistematika penulisan, yaitu sebagai berikut:

BAB I Pendahuluan

Bab ini berisi penjelasan mengenai latar belakang, rumusan masalah, tujuan dan manfaat penelitian, batasan masalah, serta sistematika penulisan.

BAB II Landasan Teori

Bab ini berisi penjelasan mengenai tinjauan pustaka dan penelitian terkait yang digunakan sebagai acuan dalam penelitian ini.

BAB III Metode Penelitian

Bab ini menjelaskan mengenai tahapan metode yang digunakan dalam penelitian ini.

BAB IV Hasil dan Pembahasan

Bab ini memuat penjelasan mengenai penerapan pembuatan program berdasarkan rancangan yang telah dibuat sebelumnya.

BAB V Penutup

Bab ini berisi kesimpulan dari penelitian yang telah dilaksanakan yang sesuai dengan rumusan masalah dan berisi saran untuk penelitian ini.