

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Pada akhir tahun 2019 dunia digemparkan oleh kemunculan virus baru yang diberi nama COVID-19. Virus ini menjadi momok yang besar karena tingkat penyebarannya yang sangat pesat, benar saja hingga tanggal 12 Maret 2020 sudah terdapat 121 negara yang terkonfirmasi terpapar virus COVID-19 [1]. Segala upaya dilakukan oleh pemerintah di berbagai negara untuk meredam persebaran virus COVID-19 tidak lain halnya dengan pemerintah Indonesia, salah satu negara yang terdampak virus COVID-19. Upaya yang dilakukan pemerintah Indonesia untuk menekan persebaran virus COVID-19 yaitu dengan kebijakan *social distancing*.

Pemerintah merupakan pihak yang berkewajiban membuat kebijakan untuk mensejahterakan masyarakatnya. Maka daripada itu perlu adanya evaluasi secara berkala terhadap kebijakan yang dikeluarkan oleh pemerintah, tapi sebelum itu perlu adanya sikap terbuka dari pemerintah untuk menerima kritik, saran, dan keluhan dari masyarakat. Pada masa pandemi virus COVID-19 pemerintah Indonesia mengeluarkan kebijakan *social distancing*, banyak opini yang tersebar di kalangan masyarakat terkait kebijakan yang dikeluarkan pemerintah baik itu opini positif maupun opini negatif. Berdasarkan hal inilah perlu dilakukannya analisis sentimen terhadap kebijakan *social distancing* yang dikeluarkan pemerintah untuk mengetahui sentimen yang beredar di masyarakat Indonesia.

Pada zaman modern ini masyarakat sudah sangat memudahkan dalam menyampaikan sebuah opini terkait suatu hal. Berdasarkan data statistik terdapat 160 juta pengguna aktif media sosial di Indonesia (*detik.com*). Dari hal tersebut dapat kita ketahui bahwa media sosial merupakan perantara yang paling banyak diminati masyarakat pada zaman ini dalam menyebarkan opini ke khalayak umum. Twitter merupakan media sosial *microblogging* yang sangat digandrungi

masyarakat Indonesia saat ini. *Microblogging* merupakan suatu layanan media sosial yang memungkinkan pengguna me-publish berupa opini, komentar, berita dalam karakter yang terbatas (kurang dari 200 karakter) [2]. *Twitter* memungkinkan pengguna untuk berbagi pesan teks singkat yang disebut tweet. *Twitter* sering kali digunakan untuk mengungkapkan emosi terkait suatu hal. Dari sini dapat kita lihat bahwa *twitter* sangatlah berpotensi untuk melihat sentimen yang ada di kalangan masyarakat terkait kebijakan social distancing yang dikeluarkan pemerintah.

Analisis sentimen sendiri sebenarnya dapat dilakukan secara manual, tetapi mengingat jumlah data tweet yang sangat banyak dan terus bertambah hal ini tidaklah mungkin dilakukan. Disinilah peran text mining diperlukan. Text mining adalah proses memeriksa sekumpulan dokumen untuk menemukan informasi baru atau membantu menjawab pernyataan penelitian tertentu (Linguamatics.com). Text mining memiliki beberapa tahap yaitu tokenizing, filtering, stemming, tagging, analyzing. Kemudian dari tahap tersebut dilakukan penggolongan pada setiap tweet, salah satunya dengan metode support vector machine (SVM).

Support vector machine (SVM) dikembangkan oleh Boser, Guyon & Vapnik pada tahun 1992, konsep SVM pada dasarnya adalah upaya pencarian nilai *hyperline* yang terbaik pemisah antara dua buah *class* dalam *input space* [3]. Kelebihan SVM dibandingkan metode yang lain terletak pada kemampuannya untuk menemukan *hyperplane* terbaik yang memisahkan dua buah class pada *feature space* yang ditunjang oleh strategi *Structural Risk Minimization* (SRM) sehingga dapat meminimalisir *error* pada data *testing*. SVM juga dibekali dengan sistem kernel yang dapat membantu dalam mengatasi permasalahan *non linier*.

Berdasarkan penelitian sebelumnya mengenai *Sentiment Analysis Of Indonesian Homosexual Tweets Using Support Vector Machine Method* didapat bahwa SVM memiliki akurasi yang baik sebesar 98.18% dengan perbandingan data latih dan data uji sebesar 90:10. Pada penelitian ini juga didapat hasil bahwa perbandingan dari jumlah data latih dan data uji mempengaruhi akurasi [4]. Pada penelitian lainnya tentang analisis sentimen menggunakan *support vector machine* dan *maximum entropy* didapatkan hasil algoritme *support vector machine*

memiliki akurasi lebih tinggi sebesar 75% dibandingkan algoritme *maximum entropy* [5]. Berdasarkan penelitian yang telah ada sebelumnya dan kelebihan dari SVM maka pada penelitian ini menggunakan algoritme *support vector machine* dalam melakukan sentimen analisis.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, maka dirumuskan masalah penelitian ini, yaitu :

1. Bagaimana mengimplementasikan algoritme *support vector machine* terhadap sentimen analisis *twitter*?
2. Bagaimana perbandingan jumlah data uji dan data latih terbaik untuk algoritme *support vector machine*?
3. Bagaimana sentimen rakyat Indonesia terhadap kebijakan *social distancing*?

1.3 Batasan Masalah

1. Data yang dipakai hanya *tweet* berbahasa Indonesia dari media sosial *twitter*.
2. *Tweet* yang diambil hanya yang mengandung kata *social distancing*, *physical distancing*, dan *#dirumahaja*. Sejak tanggal 15 Maret 2020 sampai 31 Desember 2020

1.4 Tujuan Penelitian

1. Penelitian ini memiliki tujuan untuk dapat mengklasifikasikan informasi pada *tweet* di Indonesia terkait kebijakan *social distancing* positif atau negatif menggunakan algoritme *support vector machine*
2. Mencari bagaimana jumlah perbandingan data latih dan uji yang baik untuk *support vector machine*
3. Memberikan informasi kepada khalayak umum terkait sentimen yang beredar terkait kebijakan *social distancing*

1.5 Manfaat Penelitian

1.5.1 Bagi Penulis

1. Mengaplikasikan ilmu-ilmu akademis yang didapat selama perkuliahan untuk identifikasi sentimen *twitter* terkait kebijakan *social distancing* yang dikeluarkan pemerintah Indonesia dengan algoritme *support vector machine*.
2. Memperluas wawasan khususnya dalam metode *support vector machine* khususnya dalam sentimen analisis.

1.5.2 Bagi Universitas

1. Memberikan gambaran terhadap penerapan ilmu pengetahuan yang telah diterima selama kuliah
2. Menjadi sumbangan literatur karya ilmiah dalam disiplin ilmu teknologi khususnya bidang informatika

1.6 Metode penelitian

1.6.1 Metode Pengumpulan Data

Pada tahap ini dilakukan pengumpulan data *tweet* dari *twitter* dengan menggunakan library python *twint*. Data yang diambil merupakan *tweet* yang mengandung kata *social distancing*, *physical distancing*, dan *#dirumahaja* dari tanggal 15 Maret 2020 sampai 31 Desember 2020

1.6.2 Metode Perancangan

Dalam penelitian ini, peneliti menggunakan *flowchart* untuk menunjukkan aliran di dalam program yang dibuat atau prosedur sistem secara logika dari tahap manipulasi data hingga proses klasifikasi.

1.6.3 Metode Analisis

Pada metode ini dilakukan beberapa tahapan :

1. Data Colecting

Data *tweet* akan dilakukan *crawling* menggunakan library python yaitu *twint*. Proses ini dilakukan dengan cara memasukkan keyword yang akan dicari, jumlah *tweet* yang akan diambil, dan tanggal *tweet*. *Tweet* yang telah diambil kemudain akan disimpan dalam format *csv*.

2. Preprocessing

Preprocessing merupakan tahap pembersihan data sehingga data siap digunakan untuk model klasifikasi. Terdapat beberapa tahap dalam preprocessing yaitu *cleansing*, *case folding*, *tokenisasi*, *normalization*, penghilang *stopword*, *stemming*.

3. *Feature Extraction*

Pada tahap ini, dilakukan pembobotan kata pada data yang sudah dikumpulkan sebelumnya menggunakan metode TF-IDF. Metode ini akan menentukan pembobotan kata berdasarkan frekuensi atau banyaknya kemunculan suatu kata dalam sebuah dokumen. Apabila presentase munculnya suatu kata semakin besar, maka nilainya akan semakin besar.

4. *Klasifikasi*

Pada tahap ini dilakukan klasifikasi terhadap *tweet* yang telah dilakukan *preprocessing* dan *feature extraction* sebelumnya. Metode yang digunakan dalam klasifikasi adalah SVM. SVM adalah salah satu metode *supervised learning* yang biasanya digunakan untuk klasifikasi dan regresi dengan cara mencari *hyperplane* terbaik.

1.6.4 **Metode Perancangan**

Dalam penelitian ini, peneliti menggunakan *flowchart* untuk menunjukkan aliran di dalam program yang dibuat atau prosedur sistem secara logika dari tahap manipulasi data hingga proses klasifikasi.

1.6.5 **Metode Implementasi**

Tahap implementasi merupakan penerapan dari metode analisis, seperti *data collecting*, *preprocessing*, *feature extraction*, dan *klasifikasi*.

1.6.6 **Metode Pengujian**

Model yang diperoleh pada proses pelatihan akan dilakukan pengujian *precision*, *recall*, dan *f-measure* yang akan disajikan dalam *confusion matrix* untuk mengetahui rata-rata keberhasilan dari model.

1.7 Sistematika Penelitian

Sistematika penulisan laporan Skripsi ini terdiri dari pokok-pokok permasalahan yang dibahas pada masing-masing yang akan diuraikan menjadi beberapa bagian.

BAB I PENDAHULUAN

Bab ini membahas tentang gambaran umum isi skripsi yang meliputi latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, metode penelitian, dan sistematika penelitian.

BAB II LANDASAN TEORI

Bab ini menguraikan tentang pengertian dan teori-teori yang dibutuhkan sebagai landasan atau dasar dalam penelitian yang akan dilakukan.

BAB III ANALISI DAN PERANCANGAN

Bab ini menjelaskan mengenai metode yang digunakan dalam penelitian yaitu metode pengumpulan data, metode pengembangan sistem dan lain-lainnya.

BAB IV IMPLEMENTASI DAN PEMBAHASAN

Bab ini berisi tentang paparan implementasi dan analisis hasil uji coba program dari tahap analisis, desain, implementasi desain, hasil testing dan implementasinya, berupa penjelasan teoritik, baik secara kualitatif, kuantitatif, atau secara statistik.

BAB V PENUTUP

Bab ini berisi tentang uraian mengenai kesimpulan berdasarkan penelitian yang telah dilakukan dan tentang saran-saran bagi pembaca untuk pengembangan penelitian yang telah dilakukan agar menjadi lebih baik.