

BAB I PENDAHULUAN

1.1. Latar Belakang

Dunia teknologi berkembang sangat pesat, sehingga kehidupan manusia mengalami banyak perubahan terutama setelah memasuki era digital, era dimana manusia dapat melakukan komunikasi dengan terasa dekat meskipun terpisah dengan jarak yang jauh, semua kegiatan menjadi lebih praktis dan efisien. Perkembangan teknologi memberikan pengaruh pada media dan komunikasi, hal ini dibuktikan dengan pengguna media sosial yang semakin bertambah. Salah satu media sosial yang populer yaitu Twitter. Dari data yang dikeluarkan oleh Twitter, Indonesia termasuk dalam 5 Negara dengan pengguna Twitter aktif terbesar. Pengguna Twitter aktif di Indonesia diperkirakan sebanyak 77%[1]. Twitter merupakan media sosial yang berperan besar dalam mempengaruhi beberapa aspek diantaranya dari segi sosial dan komunikasi. Pengguna Twitter diberikan kebebasan mengakses informasi dan menyampaikan opini. Opini diartikan sebagai data subjektif yang berisi tentang pandangan seseorang terhadap suatu peristiwa atau kejadian tertentu.

Kasus *covid-19* pertama kali terkonfirmasi di Indonesia pada awal bulan Maret tahun 2020, membuat Kementrian Pendidikan dan Kebudayaan atau (Kemendikbud) mengeluarkan kebijakan untuk mengubah proses pembelajaran yang awalnya tatap muka (konvensional) untuk dilaksanakan secara *online* atau daring sebagai antisipasi penyebaran *covid-19* yang semakin meluas[2]. Hampir semua lini pendidikan melakukan kegiatan pembelajaran secara daring, termasuk perguruan tinggi yang harus melaksanakan perkuliahan daring. Kebijakan perkuliahan daring cukup efektif untuk menerapkan *sosial distancing*, namun disisi lain banyak perguruan tinggi yang belum memiliki infrastruktur pembelajaran *online* yang memadai, sehingga efisiensi dalam pembelajaran menjadi berkurang. Interaksi yang kurang antara dosen dan mahasiswa juga mengakibatkan mahasiswa sulit memahami materi yang diberikan, terlebih ketika dosen yang bersangkutan belum memiliki keahlian mengembangkan teknik pembelajaran yang tepat sesuai

keadaan. Kendala lain yang sering ditemui ketika mahasiswa berasal dari wilayah terpencil sehingga kesulitan mendapatkan jaringan internet yang baik dan lancar[3].

Perkuliahan daring mempunyai sisi positif dan negatif yang beriringan, sehingga menimbulkan berbagai opini di kalangan masyarakat. Opini masyarakat biasa disampaikan lewat unggahannya pada media sosial salah satunya Twitter. Pengguna Twitter dapat memberikan opini, pendapat, maupun tanggapan terhadap kebijakan yang ditetapkan oleh pemerintah. Twitter merupakan media yang baik untuk mengambil data karena memiliki nilai akurasi kebenaran yang cukup tinggi untuk mengetahui pendapat masyarakat. Pengguna Twitter dapat mengunggah konten berisi sentimen, opini, dan *emoticon* yang dapat digunakan sebagai data untuk menganalisis kejadian atau topik tertentu. Upaya dalam menganalisis data disebut analisis sentimen.

Analisis sentimen atau *opinion mining* adalah proses untuk mempelajari dan mengolah data secara otomatis untuk memperoleh suatu informasi sentimen yang terdapat dalam kalimat opini. Analisis sentimen berguna untuk mengetahui pendapat atau kecenderungan opini oleh seseorang pada suatu masalah, kejadian, atau peristiwa yang terjadi. Analisis sentimen dapat menentukan dan mengklasifikasikan isi *dataset* yang berbentuk teks (dokumen, kalimat, paragraf) dalam sentimen positif, negatif atau netral[4]. Analisis sentimen dapat diterapkan dalam mengatasi permasalahan opini masyarakat terhadap kebijakan perkuliahan daring pada masa *covid-19* dari unggahan di media sosial Twitter yang biasa disebut *tweet*. Klasifikasi sentimen dapat mengelompokkan dan mengklasifikasikan orientasi suatu teks dalam opini positif atau negatif. Untuk melakukan analisis sentimen dibutuhkan metode yang sesuai.

Ada beberapa metode klasifikasi yang sering digunakan dalam analisis sentimen antara lain *Support Vector Machine (SVM)*, *Naïve Bayes*, *Artificial Neural Network (ANN)*. Penelitian terkait analisis sentimen diantaranya adalah penelitian oleh Luqyana dkk, pada tahun 2018 untuk menganalisis sentimen dari kolom komentar Instagram menggunakan metode *Support Vector Machine (SVM)*[5]. Kemudian penelitian Ratnawati pada tahun 2018 untuk menganalisis sentimen opini film dari media sosial Twitter menggunakan metode *Naïve Bayes* [6]. Lalu

penelitian oleh Ananto dkk, pada tahun 2019 untuk menganalisis sentimen dari data pengaduan masyarakat menggunakan metode *Artificial Neural Network (ANN)*[7]. Terakhir penelitian oleh Muthia pada tahun 2018, tentang komparasi algoritma klasifikasi menggunakan metode *Naïve Bayes* dan *Support Vector Machine(SVM)*, metode *Naïve Bayes* memperoleh hasil akurasi yang lebih unggul daripada metode *Support Vector Machine(SVM)*[8].

Metode klasifikasi teks yang paling sering digunakan adalah *Naïve Bayes*[4]. Proses klasifikasi *Naive Bayes* memanfaatkan metode probabilitas dan statistik untuk melakukan prediksi pada peluang yang berdasarkan pengalaman masa sebelumnya (*Teorema Bayes*)[6] memiliki ciri asumsi yang kuat (*naif*) bergantung pada kondisi ataupun kejadian[9]. Penelitian ini memilih menggunakan metode *Naïve Bayes*, karena merupakan metode yang sederhana, banyak digunakan pada klasifikasi teks, serta mempunyai performa yang cukup baik[10]. *Naïve Bayes* memiliki kelemahan sensitif dalam memilih fitur, apabila terlalu banyak fitur dapat meningkatkan waktu komputasi dan mempengaruhi hasil akurasi[10]. Atribut dari klasifikasi teks pada umumnya sangat besar, apabila digunakan semua akan mengurangi kinerja klasifikasi dan mempengaruhi hasil akurasi[11]. Seleksi fitur dapat melakukan pengurangan jumlah data dan melakukan identifikasi fitur yang sesuai untuk dilakukan pertimbangan pada proses pembelajaran, sehingga proses klasifikasi menjadi efisien dan efektif[12]. Metode seleksi fitur yang paling sering digunakan adalah *Information Gain*. Metode seleksi fitur *Information Gain* dipilih untuk penelitian ini karena metodenya sukses digunakan dalam klasifikasi teks[12] dan sering lebih unggul dibandingkan metode seleksi fitur lainnya[10]. Ada beberapa metode seleksi fitur lain seperti *Chi Square*, *Forward Selection*, dan *Backward Elimination*. Pada penelitian Chandani dkk, pada tahun 2015 tentang komparasi metode seleksi fitur diantaranya *Information Gain*, *Chi Square*, *Forward Selection*, dan *Backward Elimination*. Hasil dari komparasi metode seleksi fitur menunjukkan metode *Information Gain* memperoleh hasil terbaik dengan akurasi yang tertinggi mengungguli metode seleksi fitur yang lainnya[13]. *Information Gain* dapat menyeleksi dan mengurangi fitur-fitur yang kurang relevan. Data teks masih bersifat numerik dan setiap fitur dari data biasanya memiliki nilai *value* yang beragam. Metode diskritisasi atau *binning* dapat mengubah data numerik menjadi

data kategorikal untuk membantu mengurangi jumlah nilai fitur. Penelitian ini menggunakan metode diskritisasi *Equal-Width Interval* karena merupakan metode diskritisasi yang paling sederhana[14]. Pada penelitian Saleh dan Nasari, pada tahun 2018 untuk melakukan klasifikasi jurusan siswa menggunakan metode *Naïve Bayes* dan diskritisasi *Equal-Width Interval*, menunjukkan kenaikan akurasi pada hasil klasifikasi pada penerapan metode diskritisasi *Equal-Width Interval*[14]

Opini, pendapat, atau sentimen masyarakat terhadap kebijakan perkuliahan daring pada masa *covid-19* dapat memberikan masukan bagi Pemerintah untuk pengambilan keputusan atau kebijakan yang akan diambil selanjutnya. Jumlah sentimen yang sangat banyak tidak efektif apabila dilakukan secara manual, sehingga perlu dirancang sistem yang dapat melakukan klasifikasi secara otomatis untuk analisis sentimen terhadap kebijakan perkuliahan daring pada masa *covid-19* menggunakan metode *Naïve Bayes* dan metode seleksi fitur *Information Gain*.

1.2. Rumusan Masalah

Berdasarkan uraian latar belakang, maka dapat dipaparkan rumusan masalah sebagai berikut :

1.2.1. Permasalahan

Atribut dari data sentimen masyarakat terhadap kebijakan perkuliahan daring pada masa *covid-19* sangat besar, metode klasifikasi *Naïve Bayes* memiliki kelemahan sensitif dalam memilih atribut, apabila terlalu banyak atribut dapat meningkatkan waktu komputasi dan mempengaruhi hasil akurasi.

1.2.2. Usulan Solusi

Usulan solusi yang ditawarkan untuk klasifikasi sentimen masyarakat terhadap kebijakan perkuliahan daring pada masa *covid-19* menggunakan metode metode seleksi fitur *Information Gain*, untuk mempercepat waktu komputasi dan mengurangi jumlah atribut.

1.2.3. Pertanyaan Penelitian

1. Bagaimana pengaruh metode seleksi fitur *Information Gain* pada klasifikasi sentimen masyarakat terhadap kebijakan perkuliahan daring pada masa *covid-19* menggunakan metode *Naïve Bayes*?

1.3. Tujuan dan Manfaat

1.3.1. Tujuan

Tujuan penelitian untuk mengetahui pengaruh metode seleksi fitur *Information Gain* pada klasifikasi sentimen masyarakat terhadap kebijakan perkuliahan daring pada masa *covid-19* menggunakan metode *Naïve Bayes*.

1.3.2. Manfaat

Hasil penelitian ini dapat menjadi metode usulan untuk melakukan klasifikasi pada penelitian analisis sentimen yang akan datang, serta dapat menjadi rujukan untuk mengembangkan sistem dengan akurasi yang lebih tinggi.

1.4. Batasan-batasan

1. *Dataset* yang digunakan berjumlah 5.393 data dalam bentuk csv.
2. *Dataset* yang diterapkan dalam penelitian ini berasal dari jurnal *MALCOM (Indonesian Journal of Machine Learning and Computer Science)*[15], terkait sentimen masyarakat terhadap perkuliahan daring di Indonesia dari Twitter
3. Melakukan *pra-proses* teks standar tidak termasuk *convert* atau analisis *emoticon*.
4. Metode klasifikasi yang digunakan adalah *Naïve Bayes* dan metode seleksi fitur yang digunakan adalah *Information Gain*.

1.5. Sistematika Proposal

BAB I PENDAHULUAN

Pembahasan dari latar belakang, permasalahan, usulan solusi yang diajukan untuk menangani permasalahan, tujuan dan manfaat penerapan solusi yang diajukan.

BAB II TINJAUAN PUSTAKA

Teori-teori pendukung penelitian akan dijelaskan pada bab II sebagai dasar dalam pemecahan masalah.

BAB III METODE USULAN

Secara garis besar pada bab ini membahas tentang gagasan yang diusulkan mencakup langkah-langkah penelitian apa saja yang akan digunakan.

BAB IV HASIL DAN PEMBAHASAN

Menjelaskan proses uji coba sistem, analisis hasil dan evaluasi dari sistem yang telah dibuat.

BAB V PENUTUP

Membahas mengenai kesimpulan yang didapat dalam penelitian dan juga saran untuk penelitian kedepannya.