

BAB I

PENDAHULUAN

1.1 Latar Belakang

Memasuki era serba digital, internet merupakan teknologi yang paling banyak digunakan. Menurut laporan Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) sejumlah 210 juta dari total populasi penduduk Indonesia sebesar 273 juta penduduk pada tahun 2022 telah menggunakan internet (APJII, 2012). Perkembangan teknologi internet yang semakin pesat membuat teknologi informasi semakin digemari, salah satu teknologi tersebut ialah media sosial. Media sosial merupakan suatu wadah yang sering dimanfaatkan oleh masyarakat sebagai tempat untuk mencari suatu informasi, menyampaikan pendapat, serta menyampaikan isi hatinya (Rozaq et al., 2022). Salah satu media sosial yang banyak digunakan oleh masyarakat yaitu Twitter.

Data laporan *We Are Social* pada bulan Februari 2022 menyebutkan bahwa media sosial Twitter merupakan salah satu media sosial yang banyak digunakan di Indonesia sebesar 58,3% dari total jumlah penduduk (Kemp, 2022) karena Twitter merupakan sebuah jenis media sosial yang sederhana serta mudah digunakan (Giovani et al., 2020). Fitur yang ada dalam Twitter ialah fitur *tweet* dimana fitur tersebut digunakan untuk membaca serta mengirim pesan layaknya blog pada umumnya secara *real time*. Pesan teks tersebut dibatasi sebanyak 140 karakter (Anggreini et al., 2016). Ada berbagai macam berita terkini atau yang biasa disebut dengan "*trending topic*" yang sangat mudah diakses oleh masyarakat. Salah satu berita yang banyak diperbincangkan serta dibahas di Twitter ialah tentang kebijakan baru yang di gagas oleh Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi yaitu Merdeka Belajar.

Pada awal tahun 2020 Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi meluncurkan kebijakan baru yaitu Merdeka Belajar, menurutnya kebijakan tersebut merupakan sebuah langkah untuk merubah sistem pendidikan di Indonesia agar Sumber Daya Manusia (SDM) di Indonesia menjadi unggul serta

memiliki profil Pelajar Pancasila. Kebijakan tersebut memiliki empat pokok kebijakan utama antara lain penghapusan Ujian Nasional pada tahun 2021, kebijakan baru tentang penyelenggaraan Ujian Sekolah Berstandar Nasional (USBN), penerimaan peserta didik baru menerapkan sistem zonasi, serta perubahan dalam penyusunan Rencana Pelaksanaan Pembelajaran (RPP) (Kemendikbud., 2019). Kebijakan tersebut tidak lepas dari dukungan serta penolakan dari masyarakat. Masyarakat yang mendukung kebijakan tersebut menganggap bahwa kebijakan tersebut merupakan solusi baru dalam memajukan dunia pendidikan di Indonesia, sedangkan masyarakat yang menolak adanya kebijakan tersebut menganggap bahwa dengan menghapus Ujian Nasional dapat menurunkan motivasi siswa dalam belajar (Prasetyo et al., 2021). Berbagai macam upaya dukungan serta penolakan dari masyarakat dengan adanya kebijakan baru tersebut di tuangkan dalam media sosial. Komentar yang diutarakan oleh masyarakat melalui tulisan di Twitter merupakan hal yang penting untuk Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi sebagai bahan untuk mengevaluasi atau mengkaji kebijakan tersebut agar lebih baik lagi ke depannya. Dukungan serta penolakan tersebut seringkali disebut dengan sentimen. Untuk itu, cara agar mengetahui bagaimana sentimen yang diutarakan oleh masyarakat melalui media sosial yaitu dengan menerapkan teknik *Sentiment Analysis*.

Sentiment Analysis merupakan suatu proses menganalisis suatu pendapat, emosi, serta perasaan seseorang yang diungkapkan melalui teks serta mengolah suatu data dari yang tidak terstruktur menjadi data yang terstruktur. Proses *Sentiment Analysis* akan menghasilkan sentimen positif, negatif, serta netral (Nugroho et al., 2021). Untuk mengetahui komentar tersebut masuk ke dalam komentar positif, negatif, atau netral maka akan diterapkan proses klasifikasi. Berbagai macam metode yang diterapkan dalam klasifikasi *Sentiment Analysis* mempunyai nilai *accuracy* yang berbeda pula. Dari proses klasifikasi tersebut akan didapatkan hasil *accuracy* dari pengolahan menggunakan metode yang diterapkan. Beberapa macam metode yang sering diterapkan diantaranya *K-Nearest Neighbour* (K-NN), *Naïve Bayes Classifier* (NBC), *Support Vector Machine* (SVM).

Pada penelitian ini, penulis akan menerapkan metode *Naïve Bayes* dengan model *Bernoulli* untuk proses klasifikasi. *Multivariate Bernoulli* merupakan suatu model algoritma perkembangan dari *Naïve Bayes* dimana algoritma tersebut merupakan salah satu metode klasifikasi yang mempunyai probabilitas sederhana dengan menerapkan Teorema *Bayes* dimana metode ini memiliki asumsi yang kuat dari masing-masing kejadian atau kondisi (Nurhuda et al., 2016). Metode tersebut cocok digunakan pada *Sentiment Analysis* karena dapat dengan cepat serta memiliki *accuracy* yang cukup tinggi dengan jumlah dataset yang cukup besar (Pintoko & L., 2018). Penelitian sebelumnya yang dilakukan oleh Widyawati dan Susanto pada tahun 2020, penelitian tersebut membahas perbandingan kinerja antara algoritma *Multivariate Bernoulli* dengan *Multinomial Naïve Bayes*, skenario yang diterapkan yaitu pengaruh penggunaan *text preprocessing (stemming)* pada hasil *accuracy* serta waktu yang diperlukan dalam proses klasifikasi. Hasil dari penelitian tersebut *accuracy* tertinggi didapatkan dengan menerapkan metode *Bernoulli Naïve Bayes non stemming* sebesar 71.33% serta waktu tercepat dalam proses klasifikasi menerapkan metode *Multinomial Naïve Bayes non stemming* yaitu dengan waktu proses 0.12 detik. Kemudian penelitian yang dilakukan oleh Karunia pada tahun 2017, penelitian tersebut membahas tentang klasifikasi berita online menggunakan *Naïve Bayes Classifier* dengan seleksi fitur *Mutual Information*. Hasil terbaik dari metode ini adalah *accuracy* 80%, *precision* 94,28%, *recall* 79,68% dan *f-measure* 85,08% untuk *Multivariate Bernoulli* tanpa seleksi fitur. Kemudian hasil terbaik dari model klasifikasi dengan seleksi fitur dicapai pada model *Multivariate Bernoulli* dengan *accuracy* 70%, *precision* 89,11%, *recall* 69,76% dan *f-measure* 78,04% dengan tingkat efisiensi kata hingga 52% dari sebelumnya, menggunakan seleksi fitur. Sedangkan hasil *Multinomial Naïve Bayes* tanpa seleksi fitur adalah *accuracy* 41,67%, *precision* 75,68%, *recall* 41,90% dan *f-measure* 48,13%, untuk hasil model *Multinomial Naïve Bayes* dengan seleksi fitur 10% *accuracy*, 33,33% *precision*, 9,40% *recall* dan 14,35% *f-measure*. Dari beberapa penelitian tersebut membuktikan bahwa dalam proses klasifikasi menggunakan metode *Multivariate Bernoulli* mampu mengklasifikasikan dengan baik serta *accuracy* yang tinggi. Maka dari itu peneliti akan melakukan penelitian tentang *Sentiment Analysis tweet*

pada media sosial Twitter terhadap kebijakan pemerintah tentang merdeka belajar menggunakan algoritma *Multivariate Bernoulli*. Penelitian ini akan menganalisis *tweet* terhadap kebijakan merdeka belajar dengan menerapkan proses klasifikasi sentimen positif serta negatif. Pembaharuan dari penelitian sebelumnya yaitu dengan menambahkan proses Normalisasi pada *text preprocessing* dengan harapan akan meningkatkan hasil *accuracy*.

Berdasarkan latar belakang yang telah dipaparkan, maka akan dilakukan penelitian dengan judul “*Sentiment Analysis Tweet Terhadap Kebijakan Merdeka Belajar Pada Media Sosial Twitter Menggunakan Algoritma Multivariate Bernoulli*”.

1.2 Rumusan Masalah

Berdasarkan uraian pada latar belakang maka didapatkan rumusan masalah yaitu berapakah hasil *accuracy*, *precision* dan *recall* dengan menggunakan metode *Multivariate Bernoulli* pada *Sentiment Analysis Tweet* Pada Media Sosial Twitter Terhadap Kebijakan Kurikulum Merdeka Belajar?

1.3 Tujuan penelitian

Tujuan dilakukannya penelitian ini yaitu untuk mengetahui performa dari metode *Multivariate Bernoulli* pada *Sentiment Analysis Tweet* Pada Media Sosial Twitter Terhadap Kebijakan Kurikulum Merdeka Belajar.

1.4 Manfaat Penelitian

Manfaat dari penelitian ini yaitu:

1. Mengetahui performa *Multivariate Bernoulli* serta memahami tahapan dari *Sentiment Analysis* dalam melakukan klasifikasi *Tweet* pada Twitter Terhadap Kurikulum Merdeka Belajar.
2. Hasil penelitian dapat dijadikan sebagai acuan penelitian selanjutnya pada bidang *Sentiment Analysis* dengan menggunakan Algoritma *Multivariate Bernoulli* dengan Ekstraksi Fitur TF-IDF.

1.5 Batasan Penelitian

Batasan masalah pada penelitian ini yaitu:

1. Data berjumlah 1.400 data diambil pada bulan Juni – Agustus 2022.

2. “merdeka belajar” sebagai kata untuk proses pencarian data.
3. Pengambilan data dari Twitter menggunakan API (*Application Programming Interface*).
4. Hasil klasifikasi berupa sentimen positif serta sentimen negatif.
5. *Multivariate Bernoulli* sebagai algoritma proses klasifikasi.
6. *K-Fold Cross Validation* digunakan untuk validasi dengan nilai $k = 2, 5, 7, 10$.
7. *Term Frequency-Inverse Document Frequency* (TF-IDF) digunakan untuk proses pembobotan kata.
8. Menggunakan *Python* sebagai bahasa pemrograman, serta *Google Collab* sebagai *Tools*.

