

BAB I

PENDAHULUAN

1.1 Latar Belakang

Semua orang dapat dengan mudah mengekspresikan dan menyatakan pendapatnya kepada orang lain dalam kehidupan bermasyarakat. Seiring dengan perkembangan teknologi, sekarang semua orang bisa menyampaikan komentar/pemikirannya ke publik lewat jejaring sosial. Jejaring sosial adalah layanan yang memfasilitasi setiap pengguna untuk berbagi informasi dan topik dengan cakupan luas. Pendapat dapat dianalisis menggunakan analisis sentimen. Proses mengekstrak, memahami, dan memproses data tekstual dengan tujuan memperoleh informasi sentimen pada kalimat opini disebut sebagai analisis sentimen (Rozi dkk., 2012).

Salah satu platform jejaring sosial yang bisa dipakai untuk menyampaikan pendapat adalah aplikasi *twitter*. *Twitter* merupakan jejaring sosial yang sering digunakan pengguna internet. Data terbaru penggunaan *twitter* hingga saat ini mengalami peningkatan signifikan seperti yang dilansir dari *Phone Arena* pada 1 Mei 2020, mengungkapkan bahwa terdapat lonjakan pada pengguna aktif *twitter* dari 134 juta pada 2019, menjadi 166 juta pengguna dan mengalami peningkatan 24 persen (Rezeki dkk., 2020). *Twitter* dapat dijadikan salah satu sumber data pendapat/pandangan masyarakat yang bisa digunakan untuk penelitian sosial. Jumlah pengguna yang cukup besar tersebut menimbulkan banyak cuitan tentang berbagai hal.

Isu terkini yang menjadi *trending topic* pada tanggal 3 September 2022 adalah tentang kenaikan harga Bahan Bakar Minyak (BBM). BBM adalah hasil pengerjaan minyak bumi berupa cairan bahan bakar (Wahyudi dkk., 2016). Naiknya harga BBM akan menurunkan daya beli, terutama penduduk miskin dan penduduk dengan daya beli rendah, baik di perkotaan ataupun perdesaan (Muhardi, 2005). *Multiple* efek naiknya harga BBM diantaranya adalah meningkatnya biaya *overhead* perusahaan karena kenaikan biaya bahan mentah, biaya jasa angkut serta

tuntutan kenaikan upah karyawan (Arisyahidin HS, 2012). Peningkatan harga BBM ini memiliki banyak pro dan kontra, sehingga banyak pengguna *twitter* yang memberikan pendapat/pandangannya mengenai kenaikan harga BBM. Data *twitter* tersebut dapat dijadikan sebagai sumber data untuk mengevaluasi suatu kebijakan.

Tujuan dari analisis sentimen yaitu mengklasifikasi polaritas teks pada dokumen, kalimat, dan fitur serta menentukan apakah komentar/pendapat tersebut bersentimen positif, netral, atau negatif (Ramadhan & Setiawan, 2019). Analisis sentimen tentang kenaikan harga BBM perlu dilakukan untuk mengetahui respon pendapat/pandangan masyarakat khususnya pengguna *twitter* mengenai kenaikan harga BBM yang ditetapkan oleh pemerintah pada tanggal 3 September 2022. Sentimen dari masyarakat tentang kenaikan harga BBM merupakan faktor yang penting bagi pemerintah untuk menetapkan suatu kebijakan atau mengevaluasi suatu kebijakan.

Klasifikasi teks atau kategorisasi teks merupakan proses yang secara otomatis menempatkan dokumen teks ke dalam suatu kategori berdasarkan isi dari teks tersebut. Klasifikasi teks berbasis statistik, klasifikasi teks berbasis koneksi, dan klasifikasi teks berbasis aturan adalah tiga kategori utama pendekatan klasifikasi teks telah digunakan dalam penelitian. Menurut penelitian yang dilakukan oleh Yang Yiming dan Xin Liu (1999), teknik klasifikasi berbasis statistik berkinerja lebih baik daripada yang lain (Ridok & Latifah, 2015).

Salah satu metode klasifikasi berbasis statistik yang dapat digunakan untuk mengklasifikasi dokumen yaitu *multinomial naïve bayes*. Berdasarkan penelitian yang dilakukan oleh Yuyun dkk. (2021) tentang klasifikasi sentimen pemerintah terhadap penanganan Covid-19 menggunakan metode *multinomial naïve bayes* diperoleh hasil bahwa *weighted average* untuk *precision*, *recall* dan akurasi sebesar 74%. Penelitian tersebut menunjukkan bahwa akurasi metode yang diusulkan memiliki tingkatan cukup baik.

Sebagian besar studi yang menggunakan teknik klasifikasi standar mengasumsikan bahwa data *training* terdistribusi secara merata di semua kategori. Terlepas dari kenyataan bahwa ketidakseimbangan dalam jumlah data *training* sering ditemukan dalam praktiknya. Akan ada kelas minoritas dan kelas mayoritas

dalam kumpulan data yang tidak seimbang (*imbalanced class data*). Performa pengklasifikasi teks sering menurun saat dihadapkan dengan kondisi seperti ini (Ridok & Latifah, 2015).

Penyelesaian masalah distribusi data tidak seimbang ini telah diusulkan penyelesaiannya oleh Pratama dkk. (2022) yakni metode *Adaptive Synthetic* (ADASYN) yang digunakan agar proporsi kelas dalam *dataset* menjadi seimbang. Akurasi yang diperoleh dengan menggunakan metode CART yaitu sebesar 85.1% dengan data tanpa menggunakan seleksi fitur. Sedangkan akurasi yang diperoleh dengan metode *Bagging k-NN* sebesar 88% dengan data yang menggunakan seleksi fitur. Sehingga dapat disimpulkan bahwa seleksi fitur dapat meningkatkan akurasi pada klasifikasi.

Berdasarkan penelitian tersebut, peneliti akan melakukan penelitian tentang analisis sentimen mengenai kenaikan harga BBM berdasarkan *tweet* dari pengguna *twitter* pada setiap provinsi yang ada di Indonesia. *Tweet* tersebut akan diklasifikasikan menjadi tiga kategori yaitu positif, netral, atau negatif. Metode klasifikasi berbasis statistik yang digunakan yaitu *multinomial naïve bayes*. Algoritma *multinomial naïve bayes* digunakan untuk menentukan apakah *tweet* tersebut bersentimen positif, netral, atau negatif. ADASYN digunakan agar proporsi kelas pada *dataset* menjadi seimbang, sehingga model *machine learning* yang dibangun tidak cenderung mengklasifikasi suatu pendapat/komentar ke kelas mayoritas.

1.2 Rumusan Masalah

Rumusan masalah yang dikembangkan berdasarkan latar belakang yang diuraikan sebelumnya adalah sebagai berikut:

1. Bagaimana gambaran analisis sentimen pada setiap provinsi yang ada di Indonesia mengenai kenaikan harga BBM?
2. Bagaimana akurasi algoritma *multinomial naïve bayes* setelah diterapkan ADASYN?

1.3 Tujuan Penelitian

Beberapa tujuan yang ingin dicapai dalam penelitian ini adalah sebagai berikut:

1. Untuk mengetahui gambaran analisis sentimen pada setiap provinsi yang ada di Indonesia mengenai kenaikan harga BBM.
2. Untuk mengetahui akurasi algoritma *multinomial naïve bayes* setelah diterapkan ADASYN.

1.4 Manfaat Penelitian

Berdasarkan rumusan masalah dan tujuan penelitian, peneliti berharap agar penelitian ini memberikan manfaat sebagai berikut:

1.4.1 Manfaat Teoritis

- a. Memberikan gambaran serta pengetahuan dalam menganalisa sentimen pengguna *twitter* dari setiap provinsi yang ada di Indonesia mengenai kenaikan harga BBM dengan menggunakan algoritma *multinomial naïve bayes*.
- b. Penelitian ini dapat dijadikan sebagai referensi dan pengembangan pada penelitian dimasa mendatang.

1.4.2 Manfaat Praktis

- a. Bagi Peneliti
Dapat memberikan pengalaman langsung serta menambah wawasan terkait penerapan metode *multinomial naïve bayes* pada pengolahan data hasil *scraping* komentar *twitter*.
- b. Bagi Universitas
Untuk mengetahui seberapa baik mahasiswa dapat menguasai dan menerapkan ilmu yang telah dipelajari serta menjadi tolak ukur universitas dalam menilai keterampilan mahasiswa.
- c. Bagi Masyarakat
Memberikan gambaran serta pengetahuan bagi masyarakat tentang sentimen pengguna *twitter* dari setiap provinsi yang ada di Indonesia mengenai kenaikan harga BBM.

1.5 Batasan Penelitian

Batasan-batasan masalah dalam penelitian ini adalah sebagai berikut:

- a. Sentimen pengguna *twitter* dari setiap provinsi di Indonesia yang diambil dengan menggunakan *keyword* #BBMNaik, BBM Naik, dan BBM.
- b. Pengumpulan data dilakukan mulai dari tanggal 03 September 2022 sampai dengan 03 Oktober 2022. Jumlah data yang diambil pada setiap provinsi sebanyak 100 *tweet*. Total *tweet* yang digunakan pada penelitian ini sebanyak 3400. Data yang diambil merupakan *tweet* dengan Bahasa Indonesia.
- c. Klasifikasi sentimen menggunakan algoritma *multinomial naïve bayes* dengan kategori sentimen sebanyak tiga kategori yaitu sentimen positif, sentimen netral, dan sentimen negatif.
- d. *Output* dari penelitian ini adalah *wordcloud* sentimen pengguna *twitter* dari setiap provinsi yang ada di Indonesia dan akurasi algoritma *multinomial naïve bayes* setelah diterapkan ADASYN.