

BAB I

PENDAHULUAN

1.1. Latar Belakang

Bahan Bakar Minyak (BBM) merupakan komoditas yang memegang peranan penting dalam aktifitas perekonomian masyarakat. Kebijakan kenaikan harga BBM mendorong inflasi yang berpengaruh negatif terhadap pertumbuhan ekonomi. Dampak psikologis dapat terjadi dari sisi masyarakat manakala semua berekspektasi bahwa kenaikan BBM diikuti kenaikan harga sektor-sektor lainnya. Hal ini membuat produsen termasuk pedagang akan menaikkan harga untuk barang dagangannya. Penelitian telah dilakukan oleh (Latif, 2015) terhadap dampak fluktuasi harga bahan bakar terhadap suplai bahan pokok di pasar tradisional, menyimpulkan bahwa kenaikan harga BBM benar-benar sangat berdampak terhadap harga sembako, kenaikan harga sembako tergolong drastis padahal sembako adalah kebutuhan masyarakat yang harus dipenuhi setiap hari.

Dari sisi pemerintah, kebijakan kenaikan harga BBM menurunkan kredibilitas pemerintah di mata masyarakat. Menyikapi hal ini pemerintah tidak tinggal diam, berbagai kebijakan baru dibuat, seperti Bantuan Langsung Tunai BBM (BLT BBM), Bantuan Subsidi Upah (BSU), bimbingan pelaku usaha ekonomi kreatif, dan lain-lain (Kemenkeu, 2022). Presiden Joko Widodo menjamin bahwa program telah berjalan secara mudah, cepat, dan tepat sasaran (Kemenkeu, 2022). Fenomena ini menimbulkan banyak opini dari masyarakat, sehingga menimbulkan beragam sentimen. Sentimen masyarakat yang beragam dapat menjadi tolak ukur pemerintah terhadap kebijakannya juga akan menimbulkan dampak negatif yang berkepanjangan jika dibiarkan.

Analisis sentimen atau *opinion mining* adalah bidang studi yang menganalisis pendapat, sentimen, evaluasi, penilaian, sikap, dan emosi masyarakat terhadap entitas seperti produk, layanan, organisasi, individu, masalah, peristiwa, topik, dan atribut (Liu, 2012). Analisis sentimen diperlukan guna mengetahui opini masyarakat terhadap kebijakan pemerintah. Dataset adalah hal yang sangat penting dalam analisis sentimen. Analisis sentimen dilakukan terhadap ratusan ribu data *tweet* Twitter yang didapatkan melalui *TwitterAPI* untuk mengetahui kecenderungan

masyarakat Indonesia terhadap kebijakan kenaikan harga BBM, apakah ke arah positif, negatif, atau netral. Analisis sentimen terhadap *dataset* yang besar atau *big data* memiliki karakteristik yaitu ukuran data, variasi fitur, dan frekuensi kemunculan (Zikopoulos, 2011). Jumlah dan variasi dataset mempengaruhi hasil klasifikasi sentimen, semakin besar data yang digunakan pada proses *training* semakin bagus performa model klasifikasi yang dihasilkan (Tomas, 2017). Penggunaan *dataset* yang besar juga memastikan hasil klasifikasi sentimen valid karena telah mewakili suara setiap orang.

Alasan menggunakan data *tweet* Twitter karena Twitter telah menjadi salah satu platform media sosial terbesar di dunia sejak tahun 2006 (Li, 2018). Twitter mewadahi seseorang untuk berinteraksi dengan yang lain tanpa harus bertatap muka. Indonesia merupakan salah satu penyumbang terbanyak, berdasarkan data APJII, 51.5% pengguna internet di Indonesia aktif ber-sosial media. Banyak perusahaan bisnis melakukan analisis sentimen menggunakan data Twitter karena cepat dan efektif untuk memantau perasaan orang terhadap produk atau *brand* mereka (Bravo-Marquez, 2016).

Setelah *dataset tweet* terkumpul, akan dilakukan *preprocessing* untuk memudahkan dan mengoptimalkan proses klasifikasi. Lalu data akan dibagi menjadi tiga label yaitu positif, negatif, dan netral. Tahap klasifikasi *machine learning* menggunakan algoritma *Naïve Bayes Classifier (NBC)* dan *Random Forest (RF)* untuk dibandingkan berdasarkan performa ketepatan klasifikasi lalu dipilih yang terbaik. Algoritma *Naïve Bayes* adalah algoritma yang sederhana dan cocok digunakan untuk klasifikasi sentimen (Medhat, dkk., 2014), dengan bergantung pada probabilitas *Bayesian* dan asumsi *naïve* bahwa probabilitas fitur tidak bergantung satu sama lain (Hutto and Gilbert, 2015). *Random Forest* adalah algoritma *ensemble learning* untuk klasifikasi dengan menggunakan banyak pohon keputusan pada proses training dan menghasilkan prediksi akhir dari hasil prediksi kelas setiap *tree*, juga merupakan salah satu algoritma terbaik untuk klasifikasi data yang besar dengan akurasi tinggi (Genuer, 2017). Algoritma NBC dan RF merupakan metode yang terkenal dan populer digunakan untuk klasifikasi sentimen melalui *machine learning* (Bonta, dkk., 2019).

Algoritma klasifikasi *Naïve Bayes* merupakan algoritma *machine learning* yang sederhana, juga memiliki akurasi yang tinggi (Andreswari, dkk., 2019). Telah

dibuktikan pada penelitian yang dilakukan oleh (Wongkar dan Angdresey, 2019) terhadap kandidat presiden Indonesia, diketahui bahwa model *Naïve Bayes* mendapatkan akurasi terbaik dibandingkan dengan metode klasifikasi lainnya. Penelitian lain juga dilakukan oleh (Iksan, dkk., 2021) melalui analisis sentimen masyarakat Indonesia terhadap pandemi *covid-19*, didapatkan bahwa model NBC dapat mengklasifikasi data *tweet* Twitter dengan tingkat akurasi yang tinggi. Berdasarkan hasil beberapa penelitian tersebut, maka algoritma *Naïve Bayes* dipilih karena dominan memiliki akurasi tinggi pada studi kasus analisis sentimen.

Random Forest juga tergolong memiliki akurasi yang tinggi jika dibanding algoritma klasifikasi lainnya karena memiliki karakteristik *ensemble learning* (Al-Marani, 2018). Dibuktikan dengan penelitian yang telah dilakukan oleh (Karthika, dkk., 2019) terhadap sentimen penilaian kustomer pada produk-produk media sosial, dari hasil perbandingan diketahui bahwa *Random Forest* paling unggul. Penelitian lain juga dilakukan oleh (Hegde dan Padma, 2017) guna meningkatkan performa hasil klasifikasi sentimen sebelumnya terhadap ulasan produk seluler di Kanada, diketahui bahwa *Random Forest* mampu meningkatkan akurasi. Melalui beberapa penelitian tersebut, maka *Random Forest* dipilih karena dapat menghasilkan akurasi yang tinggi dengan karakteristiknya yang *ensemble learning*.

Berdasarkan referensi penelitian sebelumnya, maka penelitian ini akan menggunakan metode *Naïve Bayes Classifier* dan *Random Forest* sebagai tolak ukur ketepatan klasifikasi model pada data *tweet* Twitter. Alasan memilih dua algoritma ini selain populer dan memiliki akurasi tinggi, NBC dan RF juga merupakan algoritma yang sensitif terhadap penggunaan metode *preprocessing* (Jianqiang and Xiaolin, 2017). NBC dan RF akan digunakan sebagai metode klasifikasi sentimen masyarakat terhadap kebijakan kenaikan harga BBM di Indonesia. Harapannya hasil penelitian ini dapat menjadi tolak ukur kepuasan dan pandangan masyarakat terhadap segala kebijakan di tengah kenaikan harga BBM. Adapun tujuan menggunakan kedua metode tersebut adalah mengukur ketepatan klasifikasi dari setiap model, lalu memilih dan mengevaluasi model terbaik untuk analisis sentimen kebijakan kenaikan harga BBM. Ketepatan klasifikasi terdiri dari tingkat *accuracy*, *recall*, *precision* dan *f1-score*.

1.2. Rumusan Masalah

Permasalahan yang dikaji dalam penelitian ini adalah sebagai berikut.

1. Bagaimana mengimplementasikan masing-masing algoritma *Multinomial Naïve Bayes Classifier* dan *Random Forest* untuk mengklasifikasi sentimen masyarakat terhadap kebijakan kenaikan harga BBM di Indonesia?
2. Bagaimana hasil visualisasi data sebelum dilakukan analisis menggunakan *Naïve Bayes Classifier* dan *Random Forest*?
3. Metode mana yang lebih baik dalam mengklasifikasikan sentimen tentang kebijakan kenaikan harga BBM di Indonesia pada media sosial Twitter berdasarkan hasil analisis performansi?

1.3. Tujuan

Adapun tujuan dari penelitian ini adalah sebagai berikut.

1. Mengidentifikasi keadaan data dari hasil visualisasi sebelum dilakukan klasifikasi menggunakan algoritma *Naïve Bayes Classifier* dan *Random Forest*.
2. Mengidentifikasi kondisi model terbaik berdasarkan hasil *tuning hyperparameter* pada model klasifikasi *Naïve Bayes* dan *Random Forest*.
3. Menentukan metode yang lebih baik dalam klasifikasi sentimen kebijakan kenaikan harga BBM di Indonesia berdasarkan analisis peformansi.

1.4. Manfaat

Manfaat penelitian ini yaitu menjadi acuan informasi dan bahan evaluasi mengenai kebijakan kenaikan harga BBM di Indonesia. Memanfaatkan bidang keilmuan *data mining* dengan membandingkan metode *Naïve Bayes Classifier* dan *Random Forest* untuk klasifikasi, sehingga dapat digunakan sebagai berikut.

1. Sebagai tolak ukur kepuasan dan penilaian masyarakat terhadap kebijakan kenaikan harga BBM.
2. Menambah wawasan mengenai analisis sentimen menggunakan metode klasifikasi *Naïve Bayes Classifier* dan *Random Forest*.
3. Dapat menjadi referensi dalam pengembangan ide penelitian selanjutnya.

1.5. Batasan Masalah

Batasan masalah dalam penelitian ini adalah sebagai berikut.

1. Data yang digunakan adalah data sekunder yang didapat merupakan data *tweet* dari *platform* Twitter terkait topik kebijakan kenaikan harga BBM di Indonesia per tanggal 25 Agustus 2022 – 31 Januari 2023.
2. Menggunakan metode ekstraksi fitur *Document-term Matrix*.
3. Pendekatan distribusi pada fungsi peluang *likelihood* yang digunakan pada algoritma *Naïve Bayes Classifier (NBC)* adalah distribusi *Multinomial*.
4. Menerapkan *tuning hyperparameter* yaitu *minimum sample split* dan *minimum sample leaf* pada pengujian model *Random Forest* dan *alpha laplace smoothing* pada model *Naïve Bayes*.