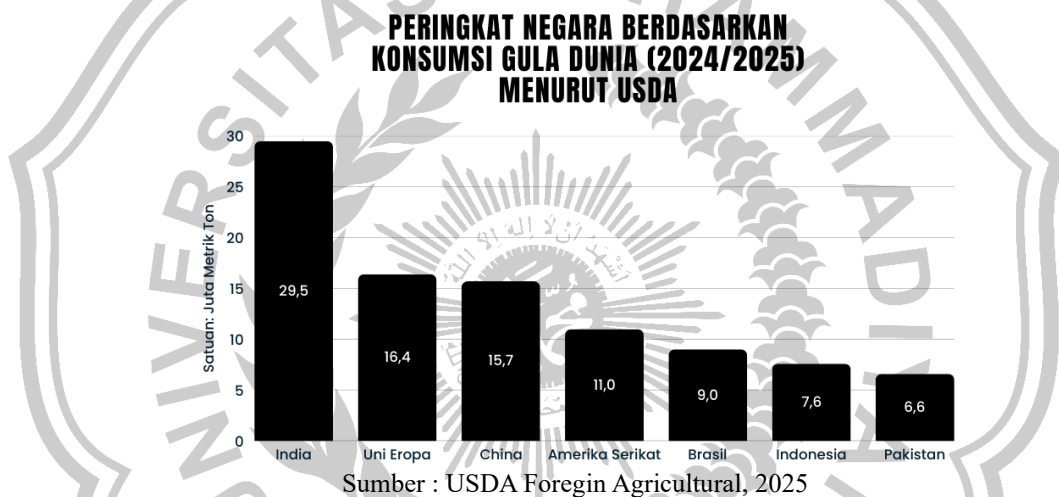


# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Menurut United States Departement of Agriculture (USDA), Indonesia menempati peringkat ke-6 dunia dalam konsumsi gula, dengan total konsumsi mencapai 7,6 juta metrik ton pada periode 2024/2025 (USDA Foreign Agricultural & Service, 2025). Faktor utama tingginya konsumsi gula ini adalah kebiasaan masyarakat Indonesia yang gemar mengonsumsi makanan dan minuman manis, terutama di kalangan anak muda (Emiliana dan Setiarini, 2024).



Sumber : USDA Foreign Agricultural, 2025

Gambar 1.1 Ranking Konsumsi Gula di Dunia Periode 2024/2025

Kementerian Kesehatan Republik Indonesia (2024) menyebutkan bahwa konsumsi gula berlebih meningkatkan risiko berbagai penyakit tidak menular (PTM) seperti diabetes, obesitas, penyakit jantung, dan kanker. Data dari International Diabetes Federation (2024) menunjukkan bahwa terdapat sekitar 20,4 juta orang di Indonesia yang hidup dengan diabetes, dengan jumlah yang terus mengalami peningkatan. Menanggapi kondisi tersebut, pemerintah memberlakukan cukai pada minuman berpemanis dalam kemasan (MBDK) sebagai upaya untuk mengurangi konsumsi gula (Direktorat Jenderal Perbendaharaan, 2025).

Kebijakan ini menimbulkan pro dan kontra di tengah masyarakat, sebagian pihak mendukung karena dinilai sebagai langkah preventif terhadap ancaman

penyakit, sedangkan sebagian lainnya menolak karena dianggap dapat berdampak negatif pada daya beli masyarakat akibat kenaikan harga, serta menurunkan pertumbuhan industri minuman. Perbedaan pendapat ini menjadikan penerapan cukai pada minuman berpemanis dalam kemasan (MBDK) sebagai isu publik yang ramai diperbincangkan, khususnya di media sosial.

Media sosial X merupakan platform yang bersifat real-time dan memiliki jangkauan luas sehingga banyak dimanfaatkan oleh masyarakat untuk mengekspresikan pandangan serta memberikan respons terhadap beragam isu publik. Berdasarkan DataReportal (2025), dengan 25,2 juta pengguna di Indonesia pada awal tahun 2025, X menjadi sumber data yang potensial dalam mengukur persepsi publik terhadap berbagai kebijakan, termasuk kebijakan cukai minuman berpemanis dalam kemasan (Darwis dkk., 2021). Banyaknya opini yang diekspresikan melalui tweet menjadikan platform ini relevan untuk dianalisis lebih lanjut guna mengetahui kecenderungan sentimen masyarakat, sehingga diperlukan pendekatan analisis sentimen yang dapat mengelompokkan opini ke dalam kategori positif maupun negatif.

Analisis sentimen adalah salah satu cabang dari pemrosesan bahasa alami (Natural Language Processing) yang berfokus pada pengenalan dan pengelompokan opini atau emosi yang terdapat dalam sebuah teks. *Multinomial Naive Bayes* merupakan metode yang umum diterapkan dalam analisis sentimen, di mana algoritma ini mengklasifikasikan teks dengan menghitung peluang kategori berdasarkan jumlah kemunculan kata. Penelitian Ningsih dkk. (2024) mengindikasikan bahwa *Multinomial Naive Bayes* memiliki performa yang lebih efisien dibandingkan algoritma lain seperti SVM, terutama dalam kecepatan pemrosesan dan efisiensi komputasi. Selain itu, Menurut Watratan dkk. (2020) juga menegaskan bahwa *Multinomial Naive Bayes* unggul dibandingkan *Support Vector Machine* dan *Logistic Regression* karena mampu mencapai performa optimal meskipun dengan data latih yang terbatas, berkat pendekatan probabilistik yang digunakan. Sementara itu, algoritma seperti *Support Vector Machine* (SVM) dan *Logistic Regression* lebih cocok untuk klasifikasi teks yang kompleks, namun memerlukan waktu komputasi yang lebih tinggi. Meskipun *Support Vector Machine* cenderung menghasilkan akurasi yang lebih tinggi, *Multinomial Naive Bayes* tetap

menjadi pilihan untuk data teks berskala besar karena mampu memberikan hasil yang stabil dengan waktu pelatihan yang relatif singkat.

Meskipun memiliki kelebihan, *Multinomial Naive Bayes* juga memiliki keterbatasan, terutama karena mengansumsikan setiap fitur (kata) sebagai unsur yang berdiri sendiri. Asumsi ini membuat model kurang mampu menangkap konteks antar kata dalam satu kalimat. Keterbatasan tersebut dapat diatasi dengan menggunakan teknik ekstraksi fitur *N-gram*, yang memungkinkan model mempertimbangkan urutan kata atau kombinasi kata yang sering muncul bersama, sehingga konteks antar kata dapat tetap diperhatikan. Penelitian Gerliandeva dkk. (2024) menunjukkan bahwa penerapan *N-gram* dari *unigram* hingga *pentagram* dapat meningkatkan akurasi *Multinomial Naive Bayes*, sedangkan penghapusan kata sinonim membantu menyeleksi dataset sehingga representasi teks menjadi lebih relevan. Selaras dengan temuan tersebut, Vincent dkk. (2024) membuktikan bahwa penggunaan *CountVectorizer* yang dikombinasikan dengan *N-gram* mampu meningkatkan akurasi klasifikasi *Multinomial Naive Bayes* secara signifikan dibandingkan metode ekstraksi fitur lainnya seperti TF-IDF.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk melakukan analisis sentimen terhadap opini masyarakat mengenai kebijakan cukai pada minuman berpemanis dalam kemasan (MBDK) dengan memanfaatkan data dari platform X. Untuk memperoleh hasil klasifikasi yang lebih akurat, digunakan pendekatan *Multinomial Naive Bayes* yang diperkuat dengan teknik ekstraksi fitur *CountVectorizer* dan *N-gram*. Selain itu, pendekatan *Lexicon-Based* diterapkan untuk memberikan label awal berdasarkan kamus sentimen yang telah disusun sebelumnya, yang hasilnya kemudian dimanfaatkan dalam proses pembelajaran model *Multinomial Naive Bayes*. Kombinasi ini diharapkan tidak hanya mengurangi dampak asumsi independensi fitur, tetapi juga meningkatkan akurasi klasifikasi, khususnya dalam menganalisis opini yang bersifat informal dan variatif pada data media sosial seperti platform X. Diharapkan hasil penelitian ini dapat memberikan gambaran mengenai persepsi masyarakat dan menjadi dasar pertimbangan dalam penerapan kebijakan cukai pada minuman berpemanis dalam kemasan (MBDK) di Indonesia.

## 1.2 Rumusan Masalah

Merujuk pada penjelasan latar belakang sebelumnya, permasalahan dalam penelitian ini dapat dirumuskan sebagai berikut:

- a. Bagaimana hasil analisis sentimen opini masyarakat terkait kebijakan cukai minuman berpemanis dalam kemasan yang diklasifikasikan menggunakan *Multinomial Naive Bayes* pada media sosial X?
- b. Bagaimana hasil evaluasi performa model klasifikasi sentimen berdasarkan akurasi, presisi, *recall*, *F1-Score*?

## 1.3 Tujuan Penelitian

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

- a. Mengetahui hasil analisis sentimen terkait kebijakan cukai minuman dalam kemasan dari penerapan *Multinomial Naive Bayes* pada media sosial X.
- b. Mengetahui nilai akurasi, presisi, *recall*, dan *F1-Score* hasil dari penerapan metode *Multinomial Naive Bayes*.

## 1.4 Manfaat Penelitian

Adapun manfaat yang diharapkan dari penelitian ini adalah sebagai berikut:

- a. Memberikan informasi dan gambaran persepsi masyarakat mengenai kebijakan cukai minuman berpemanis dalam kemasan (MBDK) berdasarkan *tweet* yang diambil dari platform X, yang dapat menjadi pertimbangan pemerintah dalam pengambilan kebijakan.
- b. Sebagai referensi untuk penelitian selanjutnya dalam bidang analisis sentimen khususnya terkait kebijakan publik menggunakan data cuitan media sosial.

## 1.5 Batasan Penelitian

Agar penelitian ini memiliki fokus yang jelas dan tidak terlalu luas, maka batasan-batasan masalah yang digunakan sebagai berikut:

- a. Penelitian ini dilakukan dalam konteks yang berkaitan dengan kebijakan penerapan cukai minuman berpemanis dalam kemasan (MBDK).

- b. Data yang digunakan berupa *tweet* dari platform media sosial X yang diperoleh menggunakan kata kunci: “cukai gula, cukai minuman, cukai minuman manis, cukai MBDK, dan cukai gula”.
- c. Proses analisis sentimen pada penelitian ini dilakukan dengan menerapkan algoritma *Multinomial Naive Bayes*.
- d. Pelabelan awal dilakukan dengan pendekatan *lexicon-based*.
- e. Ekstraksi fitur menggunakan *CountVectorizer* dengan *N-gram* yang hanya mencakup unigram dan bigram.
- f. Data berupa 500 *tweet* dari media sosial X yang telah diolah, diambil pada periode 1 Februari 2024 sampai 30 April 2025.
- g. Analisis sentimen yang dilakukan mengklasifikasikan ulasan menjadi 2 kelas, positif dan negatif.
- h. Pembagian data dilakukan menggunakan *K-Fold Cross Validation* dengan  $K = 2, 5$ , dan  $10$ .
- i. Untuk menangani ketidakseimbangan kelas pada dataset, diterapkan teknik *Borderline-SMOTE*.