BAB 2
LANDASAN TEORI

2.1 Pajak

2.1.1 Pengertian Pajak

Berdasarkan Undang-Undang Republik Indonesia Nomor 28 Tahun 2007 tentang Ketentuan Umum dan Tata Cara Perpajakan (KUP), pajak didefinisikan sebagai kontribusi wajib kepada negara yang terutang oleh orang pribadi atau badan, yang bersifat memaksa berdasarkan undang-undang, tanpa memperoleh imbalan secara langsung, dan digunakan untuk keperluan negara guna sebesar-besarnya kemakmuran rakyat (Peraturan Pemerintah RI, 2007).

Menurut Sommerfeld, Anderson, dan Brock, pajak merupakan proses pengalihan sumber daya dari sektor swasta ke sektor publik yang bersifat wajib dan bukan disebabkan oleh pelanggaran hukum. Pemungutan pajak ini didasarkan pada ketentuan perundang-undangan yang berlaku, tanpa memberikan imbalan langsung atau sebanding kepada pihak pembayar pajak (Askikarno Palalangan et al., 2019).

Definisi tersebut menegaskan bahwa pajak merupakan instrumen utama dalam pembiayaan negara yang memiliki sifat memaksa secara yuridis. Meskipun tidak ada balasan langsung yang diterima oleh pembayar pajak, hasil dari pemungutan pajak dimanfaatkan untuk membiayai berbagai program pemerintah, seperti pembangunan infrastruktur, pelayanan publik, serta penyelenggaraan fungsi negara lainnya. Dengan demikian, pajak berperan strategis dalam mendukung pertumbuhan ekonomi nasional dan pemerataan kesejahteraan masyarakat.

2.1.2 Ciri – Ciri Pajak

Pajak memiliki karakteristik yang membedakannya dari bentuk iuran atau pungutan lainnya. Adapun ciri-ciri dari pajak adalah sebagai berikut (Darmayanti, 2012) :

a) Pajak merupakan bentuk peralihan kekayaan dari masyarakat kepada negara.

b) Pemungutan pajak bersifat memaksa dan berdasarkan hukum.
c) Diperuntukkan bagi pembiayaan kepentingan umum.
d) Tidak ada kontra-prestasi langsung dari pemerintah.
e) Dikelola oleh pemerintah, baik pusat maupun daerah.
f) Digunakan untuk menutup pengeluaran negara, dan dapat dialokasikan untuk investasi publik.

2.1.3 Fungsi Pajak

Pajak memiliki dua fungsi utama yang sangat penting dalam perekonomian suatu negara (Lintang et al., 2017), yaitu:

1) fungsi anggaran (*budgetair*), yaitu pajak berperan sebagai sumber pendapatan utama bagi negara yang digunakan untuk membiayai berbagai pengeluaran, baik pengeluaran rutin seperti belanja pegawai dan operasional, maupun pengeluaran pembangunan seperti infrastruktur dan layanan publik.
2) fungsi mengatur (*regulerend*), di mana pajak dimanfaatkan sebagai instrumen kebijakan pemerintah untuk mengendalikan atau memengaruhi kondisi sosial dan ekonomi masyarakat. Melalui pengenaan tarif pajak tertentu, pemerintah dapat mendorong atau membatasi aktivitas ekonomi, misalnya memberikan insentif pajak untuk industri ramah lingkungan atau menaikkan pajak untuk barang konsumsi tertentu

2.2 Pajak Pertambahan Nilai (PPN)

Pajak Pertambahan Nilai (PPN) merupakan jenis pajak tidak langsung yang dikenakan atas transaksi penyerahan Barang Kena Pajak (BKP) dan Jasa Kena Pajak (JKP) di dalam wilayah hukum Indonesia (Daerah Pabean). PPN memiliki karakteristik utama, yaitu dikenakan atas nilai tambah yang timbul dalam setiap tahap proses produksi dan distribusi. Oleh karena itu, PPN sering disebut sebagai pajak atas konsumsi, karena pada akhirnya beban pajak ditanggung oleh konsumen (Haris Mandey, 2013). Menurut Undang-Undang Nomor 42 Tahun 2009 yang merupakan perubahan ketiga atas UU No. 8 Tahun 1983, PPN diterapkan pada penyerahan BKP/JKP di dalam negeri, impor, dan

pemanfaatan barang serta jasa dari luar negeri. Pengenaan PPN ini berlaku bagi berbagai pelaku usaha, termasuk produsen, distributor, importir, serta pemegang hak paten dan merek. Tujuan utama penerapan PPN adalah untuk meningkatkan penerimaan negara dan mendorong pertumbuhan ekonomi melalui mekanisme perpajakan yang adil dan seimbang (Peraturan Pemerintahan RI, 2009).

Penyesuaian tarif Pajak Pertambahan Nilai (PPN) diatur dalam Pasal 7 Undang-Undang Nomor 7 Tahun 2021 tentang Harmonisasi Peraturan Perpajakan (UU HPP). Dalam ketentuan tersebut, tarif PPN mengalami kenaikan dari 10% menjadi 11% yang mulai berlaku pada 1 April 2022. Selanjutnya, pemerintah diberikan kewenangan untuk menaikkan kembali tarif PPN menjadi 12% paling lambat pada 1 Januari 2025. Kebijakan ini merupakan bagian dari reformasi perpajakan yang bertujuan meningkatkan rasio pajak terhadap PDB serta memperkuat struktur penerimaan negara (Peraturan Pemerintah RI, 2021).

2.3 Analisis Sentimen

Analisis sentimen adalah suatu pendekatan komputasi yang bertujuan untuk memahami opini, penilaian, sikap, dan emosi seseorang terhadap suatu hal yang diungkapkan dalam bentuk teks (Faisol et al., 2024). Analisis sentimen adalah bagian dari ekstraksi informasi, *Natural Language Processing* (NLP), dan teks mining yang digunakan mengklasifikasikan kata, kalimat, atau dokumen berdasarkan polaritas atau perasaan yang terkandung di dalamnya (Alexander et al., 2023).

2.4 Text Mining

Text mining adalah proses menggali informasi dari data yang belum terstruktur dengan memanfaatkan teknik mining data untuk menganalisa dan mengolah data. Proses ini diawali dengan pengumpulan data, diikuti oleh prapemrosesan, sebelum data tersebut diklasifikasikan lebih lanjut (Duei Putri et al., 2022). *Text mining* secara umum didefinisikan sebagai metode eksplorasi informasi, dimana pengguna berinteraksi dengan kumpulan dokumen melalui alat analisis yang mencakup berbagai komponen data mining, salah satunya

proses kategorisasi (Amrizal, 2018). Sebagai bidang penelitian yang relative baru, *text mining* mampu memberikan solusi atas berbagai masalah seperti pemrosesan, pengorganisasian, pengelompokan, dan analisi teks yang tidak terstruktur dalam jumlah besar (Amrullah et al., 2020).

2.5 X

X adalah platform media sosial yang memungkinkan penggunanya untuk mengirim pesan singkat, yang dikenal sebagai "tweet", dengan jumlah karakter yang terbatas. X merupakan salah satu model dari media sosial yang berbentuk microblogging karena membatasi jumlah karakter setiap posting (Hadiyat Balai et al., 2017). *Microblogging* adalah konsep yang memungkinkan pengguna untuk berbagi pesan singkat. X atau X, yang diciptakan oleh Jack Dorsey, seorang pengusaha dan pengembang web asal Amerika, diluncurkan pada 21 Maret 2006. Awalnya, X membatasi jumlah karakter tweet hingga 140 karakter, namun pada November 2017, platform ini memperbarui kebijakannya dengan memperbolehkan pengguna untuk menulis hingga 280 karakter per tweet (Mega Aggriany et al., 2023). Pada akhir Juli 2023, X melakukan perubahan identitas merek dengan mengganti nama dan logo menjadi ‘X’. Rebranding ini merupakan strategi perusahaan untuk menyesuaikan diri dengan perkembangan teknologi dan perubahan kebutuhan pengguna. Selain sebagai bentuk penyesuaian, perubahan tersebut juga bertujuan untuk meningkatkan kualitas pengalaman pengguna melalui pembaruan fitur serta tampilan antarmuka yang lebih modern dan fungsional (Margesta & Sufyan Abdurrahman, 2024). Sejak diluncurkan, X telah menjadi salah satu situs yang paling sering dikunjungi di internet dan dijuluki sebagai "pesan singkat dari internet” (Zukhrufillah, 2018).

Dengan platform X, pengguna dapat mengekspresikan pendapat, berbagi pengalaman, atau menyampaikan hal-hal yang relevan dengan kehidupan mereka. Selain itu, X memudahkan pengguna untuk berbagi informasi mengenai kehidupan, pengalaman, dan aktivitas mereka secara real-time, serta berinteraksi dengan orang lain melalui percakapan yang cepat dan langsung (Hastuti et al., 2023).

2.6 *Preprocessing* Data

Preprocessing adalah proses mengola data mentah menjadi data yang terstruktur agar dapat digunakan secara optimal (Fikri et al., 2020). Data yang digunakan penelitian ini bersifat tidak terstruktur, sehingga tahap *preprocessing* menjadi langkah yang sangat diperlukan untuk memastikan data dapat diolah secara optimal (Krisdiyanto et al., 2021). Tahap ini berperan signifikan, khusunya dalam analisis sentimen, karena mampu mengurangi dimensi fitur dengan menghapus elemen – elemen yang tidak relevan dalam proses klasifikasi sentimen (Tuhuteru, 2020) (Fitriyah et al., 2020). Tahap ini memastikan data menjadi lebih terorganisir dan relevan untuk dianalisis, sehingga memudahkan analisis dan meningkatkan akurasi hasil pengolahan data. Berikut adalah tahapan preprocessing (Saputra, 2023):

a. *Cleaning*, yaitu pembersihan teks dari elemen-elemen yang tidak relevan, seperti angka, simbol, *username* atau *mention*, *hashtag*, *link*, dan *emoticon*.
b. *Case Folding*, merupakan proses mengubah seluruh kata dalam dokumen dataset menjadi huruf kecil (*lowercase*).
c. *Tokenization*, yaitu proses menyeleksi, memecah dan memotong teks dalam dokumen menjadi bagian – bagian kecil berupa kata atau *term* yang dipisahkan berdasarkan spasi (Darwis et al., 2020).
d. *Stopword Removal*, merupakan proses menghapus atau menyaring kata – kata dalam dokumen yang dianggap tidak memiliki makna penting dan tidak relevan untuk analisis (Darwis et al., 2020).
e. *Stemming*, yaitu proses menghilangkan imbuhan pada kata, seperti awalan, akhiran, atau sisipan untuk mengubahnya menjadi kata dasar sesuai dengan Kamus Besar Bahasa Indonesia (KBBI) (Darwis et al., 2020).

2.7 *Term Frequency-Inverse Document Frequency* (TF-IDF)

Term Frequency-Inverse Document Frequency (TF-IDF) adalah metode yang digunakan untuk memberikan nilai atau bobot pada kata-kata dalam sebuah dokumen dengan mempertimbangkan frekuensinya dalam dokumen tersebut serta dalam seluruh kumpulan dokumen. Teknik ini sering digunakan untuk mengidentifikasi kata-kata yang relevan, menilai kesamaan dokumen

dengan kategori tertentu, dan meningkatkan efektivitas dalam pencarian informasi. TF-IDF menggabungkan dua elemen utama, yaitu *Term Frequency (TF)*, yang mengukur seberapa sering kata muncul dalam dokumen, dan *Inverse Document Frequency (IDF)*, yang menilai pentingnya kata tersebut di seluruh kumpulan dokumen. Kedua komponen ini bersama-sama mengubah teks menjadi representasi numerik yang lebih mudah dianalisis dan diolah lebih lanjut. (Abdillah et al., 2023).

Metode TF-IDF sering digunakan dalam berbagai aplikasi, seperti pengambilan informasi, klasifikasi dokumen, mesin pencari, dan analisis teks. Dengan kemampuannya mengidentifikasi kata kunci yang relevan, TF-IDF membantu memfokuskan analisis pada informasi penting dan menjadi dasar untuk teknologi berbasis teks (Alexander et al., 2023). *Term Frequency* (TF) menggambarkan seberapa sering muncul suatu kata dalam dokumen tertentu. Semakin sering sebuah kata muncul semakin besar bobotnya, karena menunjukkan pentingnya kata tersebut dalam dokumen tersebut (Atika et al., 2022). Berikut adalah persamaan TF (Albin Pranata et al., 2024):

$$tf_{t,d} = \frac{Jumlah\ kemunculan\ kata\ t\ dalam\ dokumen\ d}{Jumlah\ total\ kata\ dalam\ dokumen\ d} \quad (2.1)$$

Keterangan:

$tf_{t,d}$: frekuensi *t* dalam dokumen *d*

Document Frequency (DF) mengukur jumlah dokumen yang mengandung kata tersebut dalam keseluruhan kumpulan dokumen. *Inverse Document Frequency* (IDF) yang merupakan kebalikan dari DF, dimana semakin sedikit dokumen yang mengandung kata tertentu, semakin tinggi nilai IDF-nya. Berikut adalah persamaan untuk nilai IDF nya:

$$idf_d = \log \frac{N}{n_t} \quad (2.2)$$

Keterangan:

idf_d : Nilai IDF dari kata *t*

N : Jumlah total dokumen dalam kumpulan dokumen

n_t : Jumlah dokumen yang mengandung kata *t*

Persamaan TF-IDF dihitung dengan mengalikan frekuensi kata dalam dokumen (TF) dengan kebalikan logaritma jumlah dokumen yang memuat kata tersebut (IDF). Oleh karena itu, persamaanya adalah:

$$tfidf_{t,d} = tf_{t,d} \times idf_d \quad (2.3)$$

2.8 *K-Nearest Neighbor* (KNN)

Algoritma *K-Nearest Neighbor* (KNN) adalah metode dalam *supervised learning* yang digunakan untuk mengklasifikasikan objek berdasarkan kedekatannya dengan data latih yang memiliki jarak terdekat dengan objek data uji (Habibah et al., 2023). Tujuan dari algoritma K-NN adalah untuk mengklasifikasikan objek baru berdasarkan atribut dan sampel latih tanpa menggunakan model yang perlu dicocokkan, melainkan mengandalkan memori. Dalam prosesnya, algoritma ini mencari K titik latihan yang paling dekat dengan titik uji, kemudian menentukan kelas objek dengan cara melakukan voting mayoritas di antara K objek tersebut. Klasifikasi dilakukan dengan pendekatan ketetanggaan untuk memprediksi kelas dari sampel uji yang baru (Gusdiana et al., 2023).

Algoritma KNN memiliki beberapa keunggulan, seperti kemampuannya yang baik dalam menangani data pelatihan dengan banyak noise dan efektivitasnya saat jumlah data pelatihan sangat besar. Namun, ada beberapa kelemahan pada KNN, seperti kebutuhan untuk menentukan nilai parameter K (jumlah tetangga terdekat), ketidakjelasan mengenai jenis jarak yang digunakan dan atribut yang perlu dipilih untuk hasil yang optimal, serta biaya komputasi yang tinggi karena setiap *query instance* memerlukan perhitungan jarak terhadap semua sampel dalam data pelatihan (Yustanti, 2012). Jarak yang digunakan dalam algoritma ini adalah jarak Euclidean, yang merupakan jenis jarak paling sering diterapkan pada data numerik. Berikut adalah persamaannya:

$$d = \sqrt{\sum_{i=1}^{n} (a_i - b_i)^2} \quad (2.4)$$

Keterangan:

d = Jarak Euclidean antara dua vector

$a_i b_i$ = Elemen ke-i dari vektor a dan b

n = Jumlah elemen (dimensi) dalam vector

Berikut ini adalah tahapan-tahapan dalam algoritma KNN:

1. Menentukan nilai K.
2. Menghitung jarak Euclidean menggunakan persamaan 2.4.
3. Mengelompokkan dan mengurutkan data berdasarkan jarak yang dihitung.
4. Menentukan klasifikasi berdasarkan mayoritas tetangga terdekat.

2.9 *Confusion Matrix*

Confusion Matrix adalah table yang digunakan untuk membandingkan hasil prediksi sistem dengan hasil sebenarnya dalam suatu klasifikasi. Table ini menunjukkan jumlah data yang diklasifikasikan dengan benar maunpun salah (Fikri et al., 2020). Metode ini digunakan untuk mengevaluasi performa model klasifikasi dengan menghitung tingkat keakuratan dan kesalahannya. Matrik ini di buat dengan membandingkan hasil prediksi sistem dengan data asli, sehingga berisi informasi nyata dan hasil perkiraan sistem. Setelah sistem melakukan klasifikasi, diperlukan pengukuran untuk mengetahui seberapa tepat hasil yang diberikan oleh sistem (Abdillah et al., 2023).

Confusion Matrix menampilkan beberapa kemungkinan dari hasil klasifikasi pada table berikut (Albin Pranata et al., 2024):

Table 2.1 *Confusion Matrix*

Kelas Aktual	Kelas Prediksi	
	Positive (+)	*Negative (-)*

Positive (+)	*True Positive (TP)*	*False Positive (FP)*
Negative (-)	*False Negative (FN)*	*True Negative (TN)*

Keterangan:

- *True Positive (TP)*: Jumlah data dalam kelas positif yang berhasil diklasifikasikan dengan benar sebagai kelas positif.
- *True Negative (TN)*: Jumlah data dalam kelas negatif yang diklasifikasikan dengan benar sebagai kelas negative.
- *False Positive (FP)*: Jumlah data dalam kelas negatif yang salah diklasifikasikan sebagai kelas positif.
- *False Negative (FN)*: Jumlah data dalam kelas positif yang salah diklasifikasikan sebagai kelas negatif.

Dalam pengukuran performa menggunakan confusion matrix, terdapat beberapa matrik yang digunakan:

1. *Accuracy* (Akurasi): mengukur efektivitas metode klasifikasi yang diterapkan

$$\frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (2.5)$$

2. *Recall* (Sensitivitas): menunjukkan rasio antara jumlah prediksi positif yang benar dengan data positif yang sebenarnya.

$$\frac{TP}{TP + FP} \times 100\% \quad (2.6)$$

3. *Precission* (Presisi): meghitung rasio antara jumlah prediksi positif yang benar dengan keseluruhan hasil prediksi positif.

$$\frac{TP}{TP + FN} \times 100\% \quad (2.7)$$

4. *F1-Score*: kombinasi rata-rata tertimbang antara *recall* dan *precision*.

$$2 \times \frac{Precision \times Recall}{Precision + Recall} \times 100\% \quad (2.8)$$

2.10 Peneltian Terdahulu

Untuk mendukung penelitian ini, diperlukan refersensi dari penelitian terdahulu yang relevan, yang disajikan dalam tabel 2.2

Table 2.2 Review Artikel

1	**Masalah**	Tata letak buku di perpustakaan yang kurang efisien dan sulit diakses.
	Metode	*K-Nearest Neighbor* (K-NN)
	Hasil Penelitian	K-NN berhasil memprediksi dan mengatur buku sesuai minat pengunjung, meningkatkan akses dan minat baca
	***Author*/Judul**	(Dollar & Zufria, 2023). Penerapan Algoritma K-Nearest Neighbor Untuk Pengelolaan Efisien Tata Letak Buku Dalam Lingkungan Perpustakaan
2	**Masalah**	Menganalisis sentimen publik terhadap aplikasi KitaLulus dari ulasan pengguna X, termasuk keluhan proses lamar kerja dan UI/UX, untuk rekomendasi peningkatan aplikasi
	Metode	*Naïve Bayes*, KNN, dan *Decision Tree*
	Hasil Penelitian	Hasil penelitian setelah dilakukan SMOTE pada tahap preprocessing, menunjukkan bahwahasil confusion matrix KNN lebih tinggi jika dibandingkan dengan algoritma Naïve Bayes dan Decision Tree, dimana KNN mendapatkan accuracy sebesar 83.33%,precision 80.36%, dan recall 71.09%.Kemudian dilanjutkan dengan Naïve Bayes yang mendapatkan accuracysebesar 65.15%,precision 60.38%, dan recall 62.56%, dan terakhir yaitu Decision Tree dengan accuracy sebesar 48.48%, precision 45.91%%, dan recall 47.23%.
	***Author*/Judul**	(Harun & Fahmi, 2024). Perbandingan Algoritma Naïve Bayes, KNN, dan Decision Tree Pada Analisis Sentimen Terhadap Ulasan Aplikasi KitaLulus
3	**Masalah**	Memahami reaksi pengguna terhadap aplikasi Chat GPT untuk meningkatkan pengalaman pengguna
	Metode	*K-Nearest Neighbor* (K-NN)

	Hasil Penelitian	KNN efektif dalam klasifikasi sentimen ulasan aplikasi chat dengan akurasi tinggi (sekitar 96.6%). Hasil analisis ini membantu pengembang memahami dan memperbaiki layanan aplikasi
	***Author*/Judul**	(Fahriza & Riza, 2023). Analisis Sentimen Pada Ulasan Aplikasi Chat Generative Pre-Trained Transformer Gpt Menggunakan Metode Klasifikasi K-Nearest Neighbor (KNN) Sistematic Literature Review
4	**Masalah**	Menganalisis sentimen terhadap Presiden Joko Widodo dengan fokus pada opini masyarakat terhadapnya selama masa jabatannya sebagai Presiden di Indonesia. Penelitian ini berfokus pada pengembangan metode *K-Nearest Neighbor* (KNN) untuk mengatasi ketidakseimbangan kelas dalam klasifikasi analisis sentimen
	Metode	*K-Nearest Neighbor* (KNN) dan SMOTE (*Synthetic Minority Over-sampling Technique*)
	Hasil Penelitian	Hasil penelitian smodel KNN dengan SMOTE menghasilkan akurasi 90%, presisi 100%, dan recall 81%, sedangkan KNN tanpa SMOTE menghasilkan akurasi 82%, presisi 82%, dan recall 98%.
	***Author*/Judul**	(Surya Firmansyah et al., 2023). Optimasi K-Nearest Neighbor Menggunakan Algoritma Smote Untuk Mengatasi Imbalance Class Pada Klasifikasi Analisis Sentimen
5	**Masalah**	Dampak pandemi Covid-19 terhadap sektor pendidikan, khususnya pembelajaran daring di Universitas Kristen Wira Wacana Sumba. Masalah utamanya adalah menganalisis sentimen atau opini mahasiswa terhadap metode pembelajaran daring selama pandemi, yang menimbulkan berbagai respons positif dan negatif.
	Metode	*K-Nearest Neighbor* (KNN)

	Hasil Penelitian	Hasil algoritma KNN dengan tujuan mencari nilai akurasi, presisi dan recall. Dari hasil penelitian diperoleh nilai sebesar 87.00% untuk accuracy dan 0.916 untuk nilai AUC.
	***Author*/Judul**	(Mara et al., n.d.). Penerapan Algoritma K-Nearest Neighbors Pada Analisis Sentimen Metode Pembelajaran Dalam Jaringan (DARING) Di Universitas Kristen Wira Wacana Sumba
6	**Masalah**	Kebijakan kenaikan PPN memunculkan kekhawatiran publik karena berpotensi meningkatkan biaya hidup dan menurunkan daya beli. Hal ini memicu beragam reaksi di media sosial, baik yang mendukung maupun yang menolak
	Metode	*Bert*
	Hasil Penelitian	Mayoritas respons publik bersifat negatif, mencerminkan kekhawatiran terhadap dampak kebijakan. Sebagian lainnya bersifat netral dengan fokus pada stabilitas ekonomi, dan sisanya menunjukkan dukungan terhadap manfaat jangka panjang. Model *Bert* menunjukkan kinerja yang baik dalam mengklasifikasi sentimen.
	***Author*/Judul**	(Imawan et al., 2025). Analisis Sentimen Publik di X Terhadap Rencana Kenaikan PPN 12% Menggunakan Bert Analysis of Public Sentiment in X Towards The 12% PPN Increase Plan Using Bert
7	**Masalah**	Dampak dari rencana kenaikan Pajak Pertambahan Nilai (PPN) menjadi 12% yang direncanakan oleh pemerintah. Kenaikan PPN ini berpotensi menyebabkan berbagai masalah ekonomi, seperti peningkatan harga barang dan jasa yang dapat mengurangi daya beli masyarakat
	Metode	*Naïve Bayes*

	Hasil Penelitian	Hasil penelitian menunjukkan bahwa dari total 468 tweet yang dianalisis, mayoritas memiliki sentimen negatif terhadap rencana kenaikan PPN, sementara sebagian kecil bersentimen positif. Hasil pengujian dan evaluasi mendapatkan nilai akurasi sebesar 83%, presisi 68,8% dan recall 78,6%.
	***Author*/Judul**	(Kristovani Siagian, 2024). Analisis Sentimen Masyarakat Indonesia Terhadap Rencana Kenaikan Ppn Menjadi 12% Di Media Sosial X Dengan Metode Naïve Bayes
8	**Masalah**	Rencana kenaikan Pajak Pertambahan Nilai (PPN) di Indonesia menjadi 12%. Kenaikan ini memicu kekhawatiran di kalangan masyarakat terkait dampaknya terhadap daya beli, harga barang dan jasa, serta potensi penurunan penjualan yang dapat melemahkan industri.
	Metode	*Decision Tree*
	Hasil Penelitian	Hasil Pengujian algoritma Decision Tree mendapatkan akurasi sebesar 81,34%, precision negatif mencapai 90,09%, sedangkan sentimen positif sebesar 75,72%, recall untuk sentimen negatif 70,42% dan untuk sentimen positif 92,25%
	***Author*/Judul**	(Diah Hardyatman & Noor Hasan, 2025). Analisis Sentimen Masyarakat Terhadap Rencana Kenaikan PPN 12% Di Indonesia Pada Media Sosial X Menggunakan Metode Decision Tree
9	**Masalah**	Mengevaluasi respon masyarakat terhadap informasi penerapan pajak pertambahan nilai atas kegiatan renovasi dan membangun rumah sendiri pada media sosial YouTube.
	Metode	*Naïve Bayes* dan *Support Vector Machine* (SVM)

	Hasil Penelitian	Respon masyarakat terhadap informasi penerapan pajak pertambahan nilai atas kegiatan renovasi dan membangun rumah sendiri beragam, dengan beberapa tidak setuju, beberapa setuju, dan ada yang tidak peduli.
	***Author*/Judul**	(Kusuma et al., 2023). Analisis sentimen masyarakat terhadap informasi penerapan PPN atas kegiatan renovasi dan membangun rumah sendiri pada media sosial youtube dengan metode svm dan naive bayes
10	**Masalah**	Ketidakseimbangan distribusi kelas data dalam analisis sentimen terhadap kenaikan PPN di media sosial, yang dapat mempengaruhi akurasi dan kinerja algoritma klasifikasi
	Metode	*Naïve Bayes* dan *Support Vector Machine* (SVM)
	Hasil Penelitian	Hasil dari penelitian ini menunjukkan bahwa algoritma SVM dengan nilai akurasi 98% sebelum dilakukan proses balancing dan 97% setelah proses balancing, sedangkan Naïve Bayes dengan nilai akurasi 97% sebelum proses balancing dan 90% setelah proses balancing. Secara umum, kedua algoritma tersebut memiliki performa yang baik dan seimbang.
	***Author*/Judul**	(Nur Ikhsan et al., 2025). Analisis Sentimen Kenaikan PPN menggunakan Algoritma Naïve Bayes dan Support Vector Machine
11	**Masalah**	Bagaimana reaksi masyarakat terhadap kenaikan tarif PPN 12% di Indonesia yang dianalisis melalui media sosial X
	Metode	*Naive Bayes dan Decision Tree*

	Hasil Penelitian	Hasil penelitian menunjukkan bahwa meskipun algoritma Decision Tree memiliki akurasi yang lebih tinggi (93,44%) dibandingkan Naive Bayes (92,68%), namun Naive Bayeslebih efisien dalam menangani dataset yang lebih besar
	***Author*/Judul**	(Adamansyah & Yudhistira, 2025). Evaluasi Opini Publik di Media Sosial X terhadap Kebijakan Pajak Pertambahan Nilai 12% di Indonesia Menggunakan Naive Bayes dan Decision Tree.
12	**Masalah**	Bagaimana persepsi publik terhadap kenaikan PPN 12% dari komentar di YouTube
	Metode	*Bert*
	Hasil Penelitian	Hasilnya menunjukkan bahwa mayoritas komentar bersentimen negatif, dan penggunaan teknologi BERT berhasil mengklasifikasi serta mengidentifikasi komentar berisi sindiran atau kiasan. Temuan ini penting untuk membantu pemerintah memahami opini masyarakat dan memperbaiki strategi komunikasi
	***Author*/Judul**	(Irmayani, 2024). Persepsi Publik Terhadap Kenaikan Ppn 12%: Pendekatan Sentimen Pada Komentar Youtube