JURNAL PENDIDIKAN INDONESIA

Jurnal Pendidikan Indonesia

p-ISSN : 2745-7141 e-ISSN : 2746-1920 Vol. 6 No. 8 Agustus 2025

Analisis Sentimen Publik X Terhadap Kenaikkan Ppn 12% pada Twitter Menggunakan Latient Dirichlet Allocation

Rizki Fajar Nulhakim, Diqy Fakhrun Shiddieq*
Universitas Garut, Indonesia
Email: 24025120028@fekon.uniga.ac.id, diqy@uniga.ac.id*

Kata Kunci:
pajak pertambahan nilai, analisis sentimen, kebijakan publik

Abstrak

Rencana kenaikan tarif Pajak Pertambahan Nilai (PPN) menjadi 12% pada tahun 2025 telah menjadi salah satu isu kebijakan publik yang hangat diperbincangkan oleh masyarakat. Kebijakan ini memunculkan berbagai reaksi, baik dukungan maupun penolakan, terutama di media sosial X (sebelumnya Twitter), yang menjadi wadah dinamis bagi masyarakat dalam menyuarakan opini. Penelitian ini bertujuan untuk menganalisis persepsi publik terhadap rencana kenaikan PPN tersebut dengan menggunakan pendekatan analisis topik dan sentimen. Metode yang digunakan adalah Latent Dirichlet Allocation (LDA), yang memungkinkan identifikasi tema-tema dominan dalam percakapan daring. Data dikumpulkan secara otomatis menggunakan Twitter Harvest dengan kata kunci "PPN 12%" dalam kurun waktu Juli hingga Oktober 2024. Data mentah kemudian melalui tahapan preprocessing sebelum dilakukan pemodelan topik. Hasil analisis menunjukkan bahwa terdapat tujuh topik utama yang muncul dari percakapan publik, meliputi isu-isu seperti daya beli masyarakat, efektivitas kebijakan pemerintah, sistem perpajakan, serta dampak ekonomi. Nilai coherence score tertinggi yang diperoleh adalah sebesar 0,88, yang menunjukkan bahwa model topik yang dihasilkan memiliki kualitas yang baik. Temuan ini memperlihatkan bahwa metode LDA mampu secara efektif menggambarkan opini dan persepsi publik terhadap isu kebijakan. Hasil penelitian ini diharapkan dapat menjadi bahan pertimbangan bagi pembuat kebijakan dalam merancang strategi yang lebih responsif terhadap aspirasi dan kekhawatiran masyarakat luas.

Keywords
value-added tax, sentiment analysis, public policy

Abstract

The plan to increase the Value Added Tax (VAT) rate to 12% by 2025 has become one of the public policy issues that is hotly discussed by the public. This policy has generated various reactions, both support and rejection, especially on social media X (formerly Twitter), which has become a dynamic forum for the public to voice their opinions. This study aims to analyze public perception of the VAT increase plan using a topic and sentiment analysis approach. The method used is Latent Dirichlet Allocation (LDA), which allows the identification of dominant

themes in online conversations. The data was collected automatically using Twitter Harvest with the keyword "12% VAT" in the period from July to October 2024. The raw data then goes through a preprocessing stage before topic modeling is carried out. The results of the analysis show that there are seven main topics that emerge from the public conversation, including issues such as people's purchasing power, the effectiveness of government policies, the tax system, and economic impacts. The highest coherence score obtained was 0.88, which indicates that the resulting topic model is of good quality. These findings show that the LDA method is able to effectively describe public opinion and perception of policy issues. The results of this research are expected to be considered for policymakers in designing strategies that are more responsive to the aspirations and concerns of the wider community.

PENDAHULUAN

Pajak Pertambahan Nilai (PPN) merupakan jenis pajak atas konsumsi yang dikenakan secara bertingkat pada setiap jalur produksi dan distribusi. Sistem PPN modern dirancang agar mekanisme pajak masukan (input tax credit) dapat mencegah efek pajak berlapis yang dapat mengganggu efisiensi produksi (Direktorat Jenderal Pajak, 2022). Dalam konteks Indonesia, struktur PPN bersifat tidak langsung, di mana beban pajak pada akhirnya ditanggung oleh konsumen, sementara pemungutannya dilakukan oleh pelaku usaha (Fathoni, 2025). Undang-Undang Nomor 42 Tahun 2009 tentang Perubahan Ketiga atas Undang-Undang Nomor 8 Tahun 1983 tentang Pajak Pertambahan Nilai Barang dan Jasa dan Pajak Penjualan atas Barang Mewah. pada bagian umum. Pajak Pertambahan Nilai adalah pajak konsumsi barang dan jasa di daerah yang dikenakan secara bertingkat di setiap jalur produksi dan distribusi(Al Ramzy et al., 2024).

Pemerintah mengesahkan Undang-Undang Nomor 7 Tahun 2021 tentang Harmonisasi Peraturan Perpajakan (UU HPP) yang menetapkan kenaikan tarif PPN dari semula 10% menjadi 11% mulai 1 April 2022, dan akan naik kembali menjadi 12% pada tahun 2025. Kebijakan ini disampaikan oleh Menteri Koordinator Bidang Perekonomian, Airlangga Hartarto, sebagai bagian dari amanat UU HPP.(Berita et al., 2025). Topik kenaikan PPN 12% dipilih karena relevansinya yang luas terhadap ekonomi masyarakat dan tingginya perhatian publik terhadap kebijakan ini. Kenaikan PPN memiliki dampak positif dan negatif terhadap kesejahteraan masyarakat. tergantung sudut pandang yang digunakan (Faisol, 2023).

Seiring dengan perkembangan teknologi. media sosial. termasuk Twitter. Twitter merupakan platform media sosial yang mudah digunakan dan memungkinkan pengguna untuk terhubung secara real-time dengan berbagai informasi global (Panggabean et al., 2025). Dengan volume data yang besar dan kemampuan penyebaran informasi secara instan, Twitter bahkan dapat

melampaui kecepatan media berita dalam melaporkan beberapa jenis peristiwa (Hou et al., 2021). Saat ini. platform media sosial dimanfaatkan secara luas untuk menyebarkan informasi. berbagi opini. serta menganalisis dampak kebijakan tersebut terhadap perekonomian dan kehidupan sehari-hari. Hal ini menjadikan Twitter sebagai platform yang efektif dalam menyuarakan berbagai pandangan publik terkait kebijakan kenaikan PPN 12%.

Twitter merupakan salah satu platform media sosial yang menyediakan berbagai cara bagi pengguna untuk berinteraksi dengan dunia (Adamansyah & Yudhistira, 2025). Pengguna dapat membuat tweet, yaitu pesan singkat yang berisi apa pun yang ingin mereka sampaikan, serta membaca berita, mengikuti akun, menyukai, atau me-retweet konten dari pengguna lain. Platform ini telah menjadi ruang bagi pertukaran informasi, opini, dan pembentukan wacana publik (Zhang et al., 2022).

Sentiment analysis merupakan studi komputasi terhadap opini, sikap, dan emosi dalam berbagai bentuk, seperti teks, gambar, dan suara, yang diarahkan pada entitas tertentu seperti isu, layanan, atau organisasi (Kumar et al., 2023). .Analisis sentimen dibutuhkan untuk mendapatkan informasi berupa teks yang dapat digunakan sebagai management reputation dan sarana evaluasi bagi perusahaan. instansi atau brand. informasi ini dapat ditemukan pada media online seperti Twitter (Imawan, Shiddieq, & Roji, 2025).

Opini dan diskusi masyarakat di Twitter terkait kenaikan PPN 12% pada 2025 menghasilkan beragam topik perbincangan. Analisis topik dengan metode Text Mining, seperti Latent Dirichlet Allocation (LDA), memungkinkan pemahaman mendalam terhadap opini, sentimen, serta dampaknya terhadap perekonomian dan daya beli masyarakat.

Latent Dirichlet Allocation (LDA) merupakan teknik dalam pemodelan topik yang bekerja dengan membentuk model probabilistik berdasarkan distribusi kata pada sekumpulan dokumen. Dalam proses ini, setiap dokumen dianggap sebagai campuran dari beberapa topik, dan setiap topik direpresentasikan sebagai distribusi probabilitas dari kata-kata. LDA digunakan untuk menghasilkan topik-topik berdasarkan kemungkinan kemunculan istilah dalam dokumen, dengan distribusi kata yang diamati pada masing-masing campuran topik dalam korpus (Onah et al., 2022a). LDA memandang setiap dokumen sebagai campuran dari berbagai topik laten, di mana setiap topik tersebut dicirikan oleh distribusi probabilistik atas kumpulan kata-kata. Dengan kata lain, dokumen terdiri atas kombinasi topik-topik tersembunyi, dan setiap topik memiliki peluang tertentu untuk muncul dalam dokumen (Guven et al., 2021).

LDA dipilih sebagai metode karena kemampuannya dalam mengidentifikasi dan mengelompokkan topik tersembunyi dalam teks tidak terstruktur. seperti cuitan di Twitter. Metode ini sangat sesuai untuk menganalisis diskusi terkait kenaikan PPN 12%. karena dapat mengolah data dalam jumlah besar secara otomatis tanpa memerlukan pelabelan awal. Penelitian terkait mengenai Latent Dirichlet Allocation (LDA) telah dilakukan

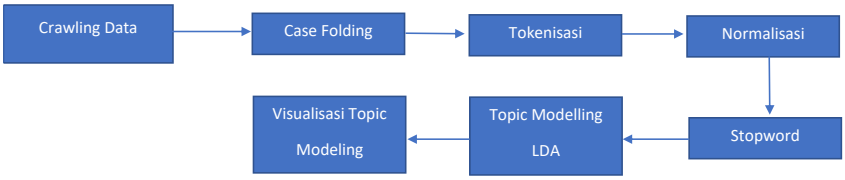
oleh (Wirasakti et al., 2020) dalam studi mereka tentang pembuatan kata kunci otomatis dalam artikel menggunakan pemodelan topik. Penelitian ini bertujuan untuk menentukan kata kunci yang relevan dalam publikasi artikel di blog dengan menerapkan model LDA. yaitu model probabilistik yang dapat mengidentifikasi beberapa topik dalam suatu teks. Data yang digunakan berasal dari berbagai blog atau artikel. kemudian diproses melalui tahap pemotongan per bagian. preprocessing. dan konversi vektor data menjadi corpus sebelum dimodelkan menggunakan LDA. Selanjutnya. dilakukan clustering dengan metode K-Means. Hasil dari pemodelan LDA menunjukkan adanya empat topik utama dengan delapan kata kunci yang memiliki probabilitas tertinggi. di antaranya: mesin (0.09375857). maksimal (0.064600445). mazda (0.10009629). varian (0.07572112). cx-8 (0.10170187). mazda (0.101048954). mobil (0.09820121). dan mazda (0.05679208).

Penelitian selanjutnya yang telah menerapkan metode Latent Dirichlet Allocation (LDA) untuk menganalisis topik terkait omnibus law berdasarkan cuitan warganet di media sosial Twitter (Luvian chisni chilmi, 2021). Penelitian ini melakukan klasterisasi topik dengan menghitung probabilitas kata terhadap topik dalam setiap iterasi. Hasil analisis menunjukkan bahwa terdapat lima topik utama yang mencerminkan berbagai perspektif masyarakat. termasuk penolakan omnibus law. kekhawatiran terhadap penyebaran Covid-19 akibat demonstrasi. dampak kebijakan terhadap lingkungan. respons humor masyarakat. serta kerusuhan saat aksi unjuk rasa. Temuan ini dapat digunakan oleh berbagai pemangku kepentingan. seperti pemerintah dan instansi terkait. sebagai bahan evaluasi dan pertimbangan dalam merumuskan kebijakan lebih lanjut.

Pengumpulan data dilakukan menggunakan Twitter Harvest, yang memungkinkan penarikan data otomatis dari Twitter secara efisien dalam periode tertentu. tool ini cocok digunakan karena menyediakan teks tidak terstruktur yang mencerminkan opini publik terkait kenaikan PPN 12%, lengkap dengan waktu dan metadata. Data yang diperoleh digunakan dalam analisis sentimen dan topik dengan metode **LDA** untuk memahami respons publik secara real-time.

METODE PENELITIAN

Penelitian ini akan dilakukan melalui sejumlah tahapan, yang dapat dilihat pada Gambar berikut.



Gambar 1. Metode

Data Crawling

Pada tahap pengambilan data, dilakukan proses crawling tweet yang relevan dengan topik penelitian berdasarkan kata kunci dan rentang waktu tertentu. Proses ini dilakukan menggunakan Tweet Harvest, sebuah *command-line tool* berbasis Playwright yang memungkinkan pengambilan tweet dari hasil pencarian Twitter. Data yang dikumpulkan disimpan dalam format CSV untuk keperluan analisis selanjutnya.

Tweet Harvest bekerja dengan memanfaatkan authorization token yang diekstrak dari akun Twitter pengguna melalui browser, serta prinsip kerja API untuk mengakses data publik. Crawling data Twitter secara umum dapat dilakukan melalui permintaan HTTP GET menggunakan kredensial aplikasi dari Twitter Developer Console, dan hasilnya dapat disimpan dalam format terstruktur untuk analisis lebih lanjut (Sohail et al., 2021). untuk melakukan crawling data. sehingga dapat mengumpulkan tweet yang mengandung kata kunci "PPN 12%".

Data Preprocessing

Setelah data terkumpul, dilakukan proses preprocessing untuk menyiapkan data sebelum pemodelan topik. Langkah awal mencakup pengambilan kolom tweet dan pembersihan data melalui *case folding*, penghapusan karakter khusus (mention, hashtag, angka, simbol), serta penghapusan *stopword* dalam bahasa Indonesia. Selain itu, dilakukan proses *tokenisasi* dan *stemming* untuk memastikan data berada dalam bentuk standar dan siap dianalisis. Tahapan ini umum dilakukan dalam studi analisis media sosial untuk meningkatkan kualitas hasil pemodelan topik seperti Latent Dirichlet Allocation (LDA) (Laureate et al., 2023). Dalam pelaksanaannya. proses preprocessing ini diterapkan dengan menggunakan bahasa pemrograman Python. Berikut secara rinci tahapan data processing LDA :

1. Case folding adalah proses standarisasi teks dengan mengubah semua huruf dalam kata menjadi huruf kecil menggunakan metode *lower* pada string di Python.
2. Tokenisasi merupakan tahap proses untuk memecah teks dokumen menjadi bentuk terkecil seperti paragraf. kalimat atau kata.
3. Normalisasi adalah proses koreksi kata agar lebih jelas dan sesuai dengan bentuk baku berdasarkan kamus.
4. Stopword removal adalah proses menghapus kata-kata umum yang tidak memiliki makna signifikan dalam analisis teks.

Topic Modelling

Tahap berikutnya dalam penelitian ini adalah membangun model Latent Dirichlet Allocation (LDA). Model ini berperan dalam mengelompokkan data serta merangkum informasi dari kumpulan data berukuran besar (Oktafiandi, 2023). LDA bekerja dengan menampilkan topik berdasarkan probabilitas kemunculan setiap kata. sehingga dapat membantu dalam menyusun dokumen secara lebih terstruktur dan sistematis(Farkhod et al., 2021).

Dalam pendekatan ini, LDA digunakan untuk menghasilkan distribusi probabilistik dari kata-kata dalam tiap topik, yang kemudian diinterpretasikan sebagai representasi sentimen positif dan negatif. Probabilitas kata-kata yang termasuk dalam masing-masing kategori dihitung untuk menentukan dominasi sentimen tertentu. Topik-topik yang dihasilkan dari data sentimen positif dan negatif memungkinkan peneliti memperoleh pemahaman yang lebih mendalam terhadap karakteristik emosi pengguna secara sistematis (Kim & Lim, 2022).

HASIL DAN PEMBAHASAN

Visualisasi

Hasil analisis topic modeling kemudian divisualisasikan menggunakan sistem visualisasi interaktif berbasis web, yaitu PyLDAvis. Visualisasi ini membantu menampilkan topik-topik yang dihasilkan oleh model LDA serta memperlihatkan hubungan antar topik secara visual. PyLDAvis digunakan untuk mengekstraksi dan menampilkan informasi dari model topik LDA dalam bentuk visualisasi interaktif berbasis web, yang memudahkan pemahaman atas prevalensi dan keterkaitan antar topik (Onah et al., 2022b). Dalam penelitian ini. penulis memanfaatkan library PyLDAvis pada platform Python. Dengan PyLDAvis. distribusi kata dalam setiap topik serta relevansinya dapat ditampilkan secara lebih jelas dan terstruktur. Hasil visualisasi ini kemudian disimpan dalam format .html. sehingga memudahkan eksplorasi hubungan antara topik serta kata-kata yang terbentuk dalam masing-masing topik.

Hasil Data Crawling

Pengambilan data dalam penelitian ini dilakukan melalui proses crawling data dari media sosial Twitter dengan memanfaatkan API Key yang disediakan

oleh Twitter. Proses crawling ini dilakukan menggunakan bahasa pemrograman Python untuk mengotomatisasi dan mempercepat pengumpulan data.

Rentang waktu pengumpulan data dimulai dari bulan Juli hingga Oktober 2024. Data yang dikumpulkan berupa opini atau tanggapan terkait topik Pajak Pertambahan Nilai (PPN) sebesar 12%. Data yang mengandung elemen visual seperti gambar. media dan link lainnya dikecualikan dari proses pengambilan untuk menjaga konsistensi analisis berbasis teks.

Data hasil crawling kemudian disimpan dalam format file .xlsx guna memudahkan proses pra-pemrosesan dan analisis lanjutan. Secara keseluruhan. jumlah data yang berhasil dikumpulkan dari Twitter sebanyak 1.900. Data ini selanjutnya digunakan sebagai dasar dalam proses analisis topik menggunakan metode Latent Dirichlet Allocation (LDA).

Hasil Data Preprocessing

Pada tahapan pra-pemrosesan (preprocessing) data. penulis menerapkan serangkaian teknik untuk membersihkan dan menyiapkan data sebelum dianalisis menggunakan algoritma Latent Dirichlet Allocation (LDA). Tahapan ini bertujuan untuk meningkatkan kualitas data teks agar lebih sesuai dan representatif dalam proses pemodelan topik.

Proses pra-pemrosesan dilakukan terhadap 1.900 data tweet yang telah diperoleh. Seluruh proses dilaksanakan menggunakan bahasa pemrograman Python pada perangkat lunak berbasis open-source. Berikut gambar hasil data preprocessing.

Tabe 1. Data Preprocessing

Tahapan	Contoh Data
Case Folding	@susipudjiastuti ya wajar aja kalo liat sikon ... gue masih sakit kepala bayangin tahun depan ba... tapi rakyatnya hanya jadi penonton dari pertun...
Tokenisasi	[susipudjiastuti, ya, wajar, aja, kalo, liat, ...] [gue, masih, sakit, kepala, bayangin, tahun, ...] [tapi, rakyatnya, hanya, jadi, penonton, dari...]
Normalisasi	[susipudjiastuti, kalang, turun, daya, beli, ...] [kepala, ppn] [rakyat, jabat, lapang, kerja, batas, phk, dim]
Stopword Removal	[susipudjiastuti, kalang, turun, daya, beli, ...]

	[rakyat, jabat, lapang, kerja, batas, phk, dim] [dana, operasional, apbn, prabowo, januari, ppn] [ppn]

Hasil Topic Modelling

Pada tahapan topic modeling menggunakan algoritma Latent Dirichlet Allocation (LDA), peneliti memanfaatkan pustaka Gensim yang dirancang untuk mengekstraksi topik-topik semantik secara otomatis dari kumpulan dokumen secara efisien menggunakan bahasa pemrograman Python. Sebelum melakukan pemodelan, peneliti terlebih dahulu menentukan jumlah topik dan kata yang mewakili setiap topik, yang dalam hal ini ditetapkan sebanyak 7 topik dan 10 kata per topik. Pemilihan jumlah ini bertujuan agar setiap topik yang terbentuk memiliki representasi yang cukup jelas dan tidak tumpang tindih dengan topik lainnya.

Untuk menilai kualitas topik yang dihasilkan, digunakan nilai coherence sebagai metrik evaluasi. Nilai ini mengukur seberapa erat keterkaitan makna antar kata dalam satu topik. Semakin tinggi nilai coherence, semakin baik dan relevan topik tersebut, karena menunjukkan kejelasan dan keterpahaman dalam konteks data. Evaluasi ini memastikan bahwa topik tidak hanya valid secara statistik, tetapi juga bermakna dan mudah dipahami, khususnya dalam analisis opini publik terkait isu PPN 12%. Berikut hasil nilai coherence.

Tabel 2. Coherence Score

Num topic	Coherence score
1	0.42
2	0.46
3	0.25
4	0.41
5	0.54
6	0.32
7	0.88

Berdasarkan Tabel 2. terlihat bahwa topik ke-7 memiliki nilai coherence tertinggi. yaitu sebesar 0.8897183641168807. Hal ini mengindikasikan bahwa tujuh topik merupakan jumlah yang paling optimal untuk memberikan representasi yang baik terhadap struktur topik dalam kumpulan data yang dianalisis. Oleh karena itu. pemodelan topik selanjutnya akan menggunakan jumlah 7 topik. guna memperoleh distribusi topik dalam dokumen serta distribusi kata dalam masing-masing topik secara lebih akurat.

Gambaran umum ini menunjukkan bahwa proses pemodelan topik menggunakan metode Latent Dirichlet Allocation (LDA) serta evaluasi melalui coherence score mampu membantu dalam menemukan jumlah topik yang paling representatif dari keseluruhan data. Nilai coherence tertinggi yang

diperoleh pada topik ke-7 memperkuat pemilihan konfigurasi tersebut sebagai dasar untuk analisis lebih lanjut.

Dalam proses evaluasi model. nilai coherence dihitung pada setiap iterasi jumlah topik untuk mengukur kualitas representasi topik yang dihasilkan. Evaluasi ini digunakan sebagai dasar dalam menentukan jumlah topik yang paling optimal. dengan mengacu pada nilai coherence tertinggi yang diperoleh. Penentuan jumlah topik yang tepat sangat berpengaruh terhadap akurasi dan kekayaan interpretasi hasil pemodelan topik.

Berikut ini merupakan beberapa contoh topik yang muncul dalam hasil pemodelan. seperti isu pajak dan PPN. program Tapera. kebijakan harga. serta persepsi publik terhadap kebijakan ekonomi pemerintah.

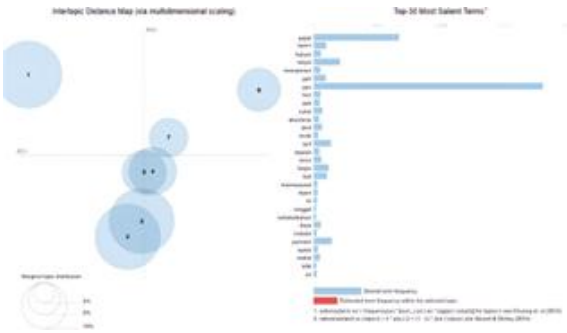
Tabel 3. Label Topik

Topik	Kata	Bobot
Kesejahteraan & Biaya Hidup	Ppn	0.33
	Rakyat	0.13
	Tapera	0.10
	Negara	0.08
	Bpjs	0.07
	Class	0.06
	Bbm	0.05
	Jabat	0.04
	Duit	0.04
	Hidup	0.04
Pajak & Layanan Publik	Ppn	0.69
	Jasa	0.16
	Tapera	0.16
	Pajak	0.16
	Rakyat	0.15
	Kerja	0.11
	Asuransi	0.08
	Debit	0.07
	Barang	0.07
	Wajib	0.07
Pajak Properti & Konsumsi Rumah Tangga	Ppn	0.89
	Pajak	0.24
	Tarif	0.14
	Bangun	0.12
	Persen	0.10
	Rumah	0.10
	Tapera	0.08
	Harga	0.07
	Makan	0.06
	Kms	0.05
Kebijakan Pemerintah & Dampaknya ke Rakyat	Ppn	0.51
	Rakyat	0.23
	Pajak	0.22
	Bayar	0.19

	Tapera	0.12
	Jokowi	0.06
	Prabowo	0.06
	Gaji	0.05
	Mahal	0.05
Harga Konsumen & Daya Beli	Ppn	0.095
	Beli	0.16
	Persen	0.14
	Perintah	0.12
	Daya	0.10
	Harga	0.10
	Masyarakat	0.08
	Barang	0.07
	Turun	0.07
	Pajak	0.07
Pajak & Isu Politik Terkait Ekonomi	Ppn	0.68
	Pajak	0.16
	Persen	0.13
	Perintah	0.10
	Rakyat	0.08
	Prabowo	0.07
	Sri	0.07
	Ikn	0.06
	Gaji	0.06
	Mahal	0.06
Sistem Perpajakan & Regulasi Keuangan	Pajak	0.107
	Ppn	0.69
	Pph	0.22
	Cukai	0.20
	Hukum	0.18
	Bea	0.17
	Manajemen	0.17
	Ppb	0.15
	Akuntansi	0.13
	Daerah	0.12

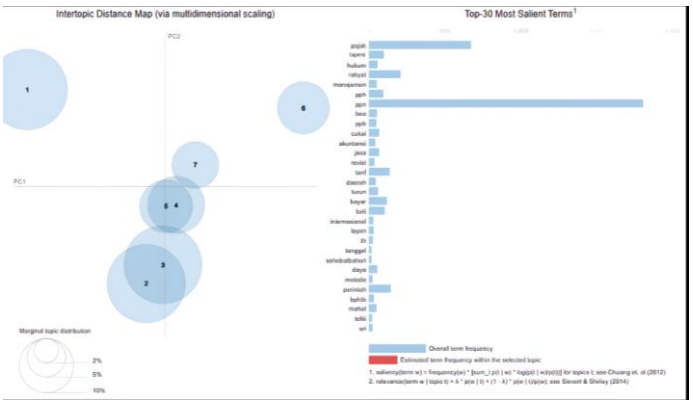
Hasil Visualisasi

Setelah proses pemodelan topik dilakukan menggunakan algoritma Latent Dirichlet Allocation (LDA), langkah selanjutnya adalah melakukan visualisasi hasil model dengan jumlah topik sebanyak tujuh. Visualisasi ini dilakukan menggunakan library pyLDAvis untuk mempermudah pemahaman terhadap distribusi kata pada masing-masing topik yang telah terbentuk dari proses pemodelan tersebut.



Gambar 1. Hasil Visualisasi

Visualisasi hasil pemodelan topik menggunakan metode Latent Dirichlet Allocation (LDA) menunjukkan bahwa jumlah topik sebanyak 7 topik berhasil mengelompokkan berbagai tema yang dibahas dalam kumpulan data tweet menjadi beberapa bagian yang berbeda. Visualisasi ini terdiri dari dua bagian utama. Berikut gambar hasil visualisasi menggunakan library pyLDavis.



Gambar 2. Visualisasi

Visualisasi hasil analisis topik menggunakan pyLDAvis menampilkan dua bagian utama. Di sisi kiri, lingkaran-lingkaran merepresentasikan topik yang terbentuk, di mana jarak antar lingkaran menunjukkan kemiripan semantic semakin dekat, semakin mirip. Ukuran lingkaran mencerminkan proporsi dominasi topik dalam keseluruhan dokumen. Beberapa topik tampak saling berdekatan dan tumpang tindih (misalnya topik 2–5), menandakan keterkaitan tema, sedangkan topik 1 dan 6 lebih terpisah, mencerminkan topik yang lebih spesifik.

Sisi kanan menampilkan diagram batang horizontal yang memperlihatkan 30 kata paling relevan dan sering muncul dalam topik, seperti *ppn*, *pajak*, *rakyat*, *tarif*, dan *cukai*. Kata-kata ini menggambarkan fokus diskusi publik terkait kebijakan pajak, peran negara, dan beban ekonomi masyarakat (Klikpajak, 2025).

Secara keseluruhan, visualisasi ini membantu menggambarkan struktur dan hubungan antar topik dengan jelas, serta menegaskan bahwa isu perpajakan,

khususnya PPN, menjadi perhatian utama dalam percakapan publik di media sosial.

KESIMPULAN

Berdasarkan hasil analisis yang telah dilakukan. penerapan metode Latent Dirichlet Allocation (LDA) terhadap 1900 data tweet yang membahas berbagai opini publik terkait PPN. pajak. kebijakan pemerintah. dan isu ekonomi-sosial menunjukkan hasil yang cukup baik. Model topik yang dihasilkan mampu mengelompokkan isu-isu utama ke dalam topik-topik yang relevan. seperti kebijakan perpajakan. daya beli masyarakat. layanan publik. dan respons terhadap program pemerintah.

Dari evaluasi model dengan menggunakan coherence score. diperoleh bahwa pemodelan dengan jumlah 7 topik memberikan hasil terbaik dengan nilai coherence tertinggi sebesar 0.88. Ini menunjukkan bahwa konfigurasi tersebut paling optimal dalam menangkap variasi tema yang muncul dalam data.

Temuan ini menunjukkan bahwa LDA dapat digunakan secara efektif untuk mengidentifikasi dan memahami pola percakapan serta opini masyarakat di media sosial. khususnya Twitter. terkait kebijakan ekonomi. Ke depannya. efektivitas pemodelan dapat ditingkatkan dengan menggabungkan metode LDA dengan pendekatan lain seperti TF-IDF. serta mempertimbangkan evaluasi tambahan seperti confusion matrix atau metode klasifikasi lanjutan guna mengukur akurasi dan kekonsistenan hasil topik.

REFERERENSI

Adamansyah, D., & Yudhistira, R. (2025). Evaluasi Opini Publik di Media Sosial X terhadap Kebijakan PPN 12% di Indonesia Menggunakan Naive Bayes dan Decision Tree. JPTI. Diakses dari jpti.journals.id.

Al Ramzy, Asnawi Asnawi, Wahyudin Wahyudin, & Dini Vientiany. (2024). Pajak Pertambahan Nilai Barang dan Jasa Serta Pajak Penjualan Atas Barang Mewah. *Jurnal Pajak Dan Analisis Ekonomi Syariah*, *1*(3), 130–136. https://doi.org/10.61132/jpaes.v1i3.260

Berita, C., Bpk, J., Provinsi, P., Barat, S., & Panji, A. (2025). *Pemerintah resmi kenakan pajak pertambahan nilai 12 persen per 1 januari 2025*. 1–9.

Direktorat Jenderal Pajak (2022). Di Balik Kenaikan Tarif PPN: UU HPP dan Penerapannya ke 12%. Diakses dari pajak.go.id.

Faisol (2023). Perspektif Netizen terhadap Kenaikan Tarif PPN (Naik 11% April 2022 dan rencana 12% Januari 2025). Jurnal Akuntansi Aktual, Universitas Muhammadiyah Malang. Diakses dari ejournal.umm.ac.id.

Farkhod, A., Abdusalomov, A., Makhmudov, F., & Cho, Y. I. (2021). Lda-based topic modeling sentiment analysis using topic/document/sentence (Tds) model. *Applied Sciences (Switzerland)*, *11*(23). https://doi.org/10.3390/app112311091

Fathoni, M. I. (2025). *Indonesia ' s Value Added Tax (VAT) Reform : Fiscal Benefits and Sectoral Impacts*. *6*(1).

Guven, Z. A., Diri, B., & Cakaloglu, T. (2021). *n-stage Latent Dirichlet Allocation: A Novel Approach for LDA*. https://doi.org/10.1109/UBMK.2019.8907050

Hou, Q., Han, M., Qu, F., & He, J. S. (2021). Understanding social media beyond text: A reliable practice on Twitter. *Computational Social Networks*, *8*(1). https://doi.org/10.1186/s40649-021-00088-x

Imawan, M. R., Shiddieq, D. M., & Roji, A. (2025). Analisis Sentimen Publik di X Terhadap Rencana Kenaikan PPN 12% Menggunakan BERT. Diakses dari researchgate.net.

Kim, J. Y., & Lim, C. K. (2022). The Use of Sentiment Analysis and Latent Dirichlet Allocation Topic-Modeling (LDA) on Web Novel Content Quality Factor. *Journal of System and Management Sciences*, *12*(2), 242–257. https://doi.org/10.33168/JSMS.2022.0211

Klikpajak (2025). Dampak Kenaikan Tarif PPN 12 Persen pada Barang Mewah. Diakses dari klikpajak.id.

Kumar, S., Roy, P. P., Dogra, D. P., & Kim, B.-G. (2023). *A Comprehensive Review on Sentiment Analysis: Tasks, Approaches and Applications*. 1–29. http://arxiv.org/abs/2311.11250

Laureate, C. D. P., Buntine, W., & Linger, H. (2023). A systematic review of the use of topic models for short text social media analysis. In *Artificial Intelligence Review* (Vol. 56, Issue 12). Springer Netherlands. https://doi.org/10.1007/s10462-023-10471-x

Luvian chisni chilmi, M. (2021). Latent Dirichlet Allocation (LDA) Untuk Mengetahui Topik Pembicaraan Warganet Twitter Tentang Omnibus Law. *Universitas Islam Negeri Syarif Hidayatullah*, 1–131. https://repository.uinjkt.ac.id/dspace/handle/123456789/56724

Oktafiandi, H. (2023). Implementasi LDA untuk Pengelompokan Topik Twitter Bertagar# Mypertamina. *Jurnal Ekonomi Dan Teknik Informatika*, *11*(1), 10–16. https://www.e-journal.polsa.ac.id/index.php/jneti/article/view/222%0Ahttps://www.e-journal.polsa.ac.id/index.php/jneti/article/download/222/154

Onah, D. F. O., Pang, E. L. L., & El-Haj, M. (2022b). A Data-driven Latent Semantic Analysis for Automatic Text Summarization using LDA Topic Modelling. *Proceedings - 2022 IEEE International Conference on Big Data, Big Data 2022*, 2771–2780. https://doi.org/10.1109/BigData55660.2022.10020259

Panggabean, S., et al. (2025). Analisis Sentimen Pengguna X terhadap Kebijakan PPN 12% Menggunakan Naive Bayes. Pustaka Data. Diakses dari jurnal.pustakagalerimandiri.co.id.

Sohail, S. S., Khan, M. M., Arsalan, M., Khan, A., Siddiqui, J., Hasan, S. H., & Alam, M. A. (2021). *Crawling Twitter data through API: A technical/legal perspective*. http://arxiv.org/abs/2105.10724

Wirasakti, L. A., Permadi, R., Hartanto, A. D., & Hartatik, H. (2020). Pembuatan Kata Kunci Otomatis Dalam Artikel Dengan Pemodelan Topik. *Jurnal Media Informatika Budidarma*, *4*(1), 27. https://doi.org/10.30865/mib.v4i1.1707

Zhang, Y., Shirakawa, M., Wang, Y., Li, Z., & Hara, T. (2022). Twitter-aided decision making: a review of recent developments. *Applied Intelligence*, *52*(12), 13839–13854. https://doi.org/10.1007/s10489-022-03241-9