



## Analisis Sentimen Publik Berbasis KNN terhadap Kinerja Purbaya dan Sri Mulyani

Krisna Diva<sup>\*1</sup>, Ari Anshari Yuliansyah Siregar<sup>2</sup>, M Azwan<sup>3</sup>, Irfan Nainggolan<sup>4</sup>,  
Rian Farta Wijaya<sup>5</sup>, Zulham Sitorus<sup>6</sup>

<sup>1,2,3,4,5,6</sup> Pascasarjana, Magister Teknologi Informasi, Universitas Pembangunan Panca Budi, Medan, Indonesia

<sup>1,\*</sup>[krisnadiva0412@gmail.com](mailto:krisnadiva0412@gmail.com), <sup>2</sup>[arianshari32@gmail.com](mailto:arianshari32@gmail.com), <sup>3</sup>[azwanatlantica48@gmail.com](mailto:azwanatlantica48@gmail.com),  
<sup>4</sup>[irfannainggolan05@gmail.com](mailto:irfannainggolan05@gmail.com), <sup>5</sup>[rianfartawijaya@gmail.com](mailto:rianfartawijaya@gmail.com), <sup>6</sup>[zulhamsitorus@gmail.com](mailto:zulhamsitorus@gmail.com)

### Abstract

Several parts of public are concerned about the reshuffling of the Financial Minister. It is because the position plays a crucial role in maintaining the country's economic stability. In Merah Putih Cabinet, Purbaya Yudhi Sadewa serves as the new Financial Minister where previously it was handled by Sri Mulyani. Thus, this reshuffling was not convincing enough to increase the public trust. It is because this position is a vital role in managing the country financial. So, it is necessary to conduct a performance analysis of these public officials. Based on this issue, the analysis of public perception for their first month of their tenure is very important. This study uses the K-Nearest Neighbor (KNN) algorithm to test public sentiment toward the thoughts on social media platform X during the first month of the second term of both figures. This dataset includes 2,071 tweets from Sri Mulyani (October 21–November 20, 2024) and 2,960 tweets from Purbaya Yudhi Sadewa (September 8–October 7, 2025). After preprocessing and data labeling with IndoBERT, the sentiment distribution for Purbaya is 57.80% positive and 42.20% negative while for Sri Mulyani it is 46.74% positive and 53.26% negative. TF-IDF will then be used for feature extraction, while KNN with  $K=5$  and the cosine similarity distance metric will be used for classification. According to the evaluation results, the KNN model for Sri Mulyani has an accuracy of 77% with a precision of 78%, recall of 77% and an F1 score of 77% while the KNN model for Purbaya has an accuracy of 80% with a precision of 80%, recall of 78% and an F1 score of 79%. Thus, it can be concluded from these evaluation results that Purbaya is more trusted by the public as the new Minister of Finance.

**Keywords:** Sentiment Analysis, K-Neighbor, IndoBERT, TF-IDF, Indonesian Minister of Finance

### Abstrak

Sebagian masyarakat khawatir tentang perubahan Menteri Keuangan karena memiliki peran penting dalam menjaga stabilitas ekonomi negara. Pada Kabinet Merah Putih, Purbaya Yudhi Sadewa menggantikan Sri Mulyani sebagai Menteri Keuangan. Namun, pergantian Menteri belum tentu dapat menambah kepercayaan masyarakat. Hal tersebut karena posisi ini merupakan posisi yang vital dalam mengelola keuangan negara sehingga perlu adanya analisis kinerja terhadap pemangku jabatan. Berdasarkan permasalahan tersebut, melakukan analisis persepsi publik selama bulan pertama menjabat pada kedua pejabat ini sangat penting. Studi ini menggunakan algoritma K-Nearest Neighbor (KNN) untuk menguji sentimen publik terhadap pemikiran di platform media sosial X selama bulan pertama masa jabatan kedua tokoh tersebut. Dataset ini mencakup 2.071 cuitan dari Sri Mulyani (21 Oktober–20 November 2024) dan 2.960 cuitan dari Purbaya Yudhi Sadewa (8 September–7 Oktober 2025). Setelah pra-proses dan pelabelan data dengan IndoBERT, distribusi sentimen untuk Purbaya adalah 57,80% positif dan 42,20% negatif, sedangkan untuk Sri Mulyani adalah 46,74% positif dan 53,26% negatif. TF-IDF kemudian akan digunakan untuk ekstraksi fitur, sementara KNN dengan  $K=5$  dan metrik jarak kesamaan kosinus akan digunakan untuk klasifikasi. Menurut hasil evaluasi, model KNN Sri Mulyani memiliki akurasi 77% dengan presisi

© 2025 Author



78%, recall 77%, dan skor F1 77%, sedangkan model KNN Purbaya memiliki akurasi 80% dengan presisi 80%, recall 78%, dan skor F1 79%. Dengan demikian, dapat disimpulkan dari hasil evaluasi ini bahwa Purbaya lebih dipercaya oleh masyarakat sebagai Menteri Keuangan yang baru.

Kata kunci: Analisis Sentimen, K-Neighbor, IndoBERT, TF-IDF, Menteri Keuangan Indonesia

## 1. Pendahuluan

Indonesia sebagai negara demokrasi dengan sistem pemerintahan presidensial menempatkan Menteri Keuangan sebagai salah satu posisi strategis dalam menjaga stabilitas ekonomi dan fiskal nasional. Setiap pergantian Menteri Keuangan selalu menjadi sorotan publik karena kebijakannya berdampak langsung pada kondisi perekonomian masyarakat. Dalam Kabinet Merah Putih di bawah kepemimpinan Presiden Prabowo Subianto, terjadi pergantian Menteri Keuangan dari Sri Mulyani kepada Purbaya Yudhi Sadewa yang dilantik pada 8 September 2025 sebagai bagian dari *reshuffle* kabinet [1]. Purbaya yang sebelumnya menjabat sebagai Kepala Lembaga Penjamin Simpanan menggantikan Sri Mulyani yang telah menjabat sejak Oktober 2024. Pergantian ini menimbulkan berbagai ekspektasi dan kekhawatiran di kalangan masyarakat terkait keberlanjutan kebijakan fiskal dan arah ekonomi nasional.

Persepsi publik terhadap seorang pemimpin, khususnya pada bulan pertama masa jabatan, menjadi indikator penting dalam mengukur tingkat kepercayaan dan penerimaan masyarakat. Periode awal ini sering kali menjadi masa krusial dimana kebijakan-kebijakan strategis mulai dirumuskan dan langkah-langkah konkret mulai diambil, sehingga ketidakpastian yang muncul dapat mempengaruhi legitimasi dan efektivitas kepemimpinan ke depannya [2]. Di era digital saat ini, media sosial seperti *platform X* (sebelumnya *Twitter*) telah menjadi ruang publik yang sangat aktif untuk mengekspresikan pendapat, kritik, dan dukungan terhadap berbagai isu [3]. *Platform X* memungkinkan masyarakat untuk menyampaikan opini secara langsung, spontan, dan masif terhadap kinerja pejabat publik, menghasilkan data yang kaya akan sentimen masyarakat yang perlu dianalisis secara objektif dan sistematis. Namun, metode komputer menggunakan analisis sentimen diperlukan untuk mengekstraksi dan mengkategorikan opini publik secara otomatis karena jumlah data yang sangat besar dan sifatnya yang tidak terstruktur membuat analisis manual tidak efektif [4].

Menurut Rhamadanti, dkk, (2024), keunggulan dari algoritma KNN adalah mampu mengatasi ketidakseimbangan kelas dalam dataset karena menggunakan mayoritas tetangga terdekat untuk pengklasifikasian. Hasil evaluasi pada penelitian ini juga memberikan akurasi sebesar 86-87%. Sedangkan menurut Nurhidayat dan Dewi, (2023), penelitian ini menggunakan KNN yang dilakukan ekstraksi fitur *Ngram* menjadi *Unigram*, *Bigram*, *Trigram* dan gabungan ketiganya. Dari hasil evaluasi model penelitian, diperoleh akurasi sebesar 98,6% pada pengujian aspek kemas dengan data seimbang menggunakan *Trigram*.

Metode memahami, mengekstraksi, dan menganalisis materi tekstual secara mekanis untuk menentukan informasi sentimen yang terkandung dalam suatu opini dikenal sebagai analisis sentimen. [5]. Dalam konteks media sosial, analisis sentimen dapat membantu memahami bagaimana masyarakat merespons suatu isu atau tokoh tertentu. Mengingat data dari *platform X* tidak memiliki label sentimen secara otomatis, diperlukan metode pelabelan yang akurat. Untuk mengatasi tantangan pelabelan data dalam skala besar, model berbasis *transformer* seperti *IndoBERT* yang dikembangkan khusus untuk bahasa Indonesia mampu menangkap makna teks secara lebih efektif dan menunjukkan keunggulan dalam membedakan sentimen positif, negatif, atau netral pada data *Twitter* berbahasa Indonesia [6].

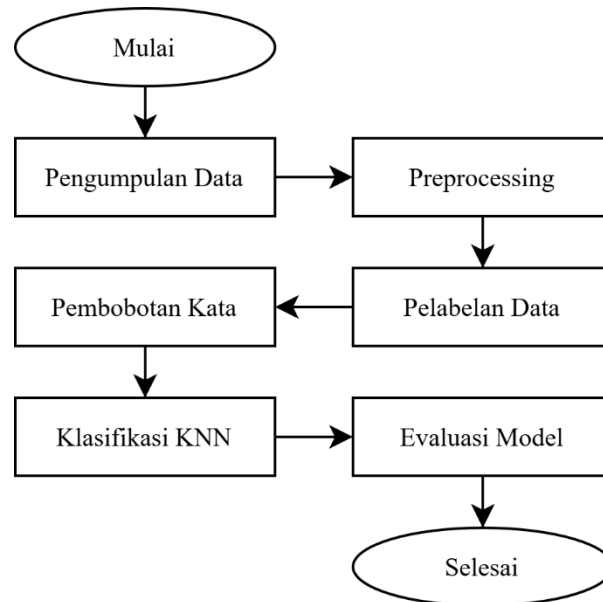
Setelah data diberi label sentimen, tahap selanjutnya adalah membangun model yang dapat memprediksi sentimen pada data baru secara otomatis. *K-Nearest Neighbor* (KNN) adalah salah satu pendekatan yang dapat digunakan; pendekatan ini dapat menghasilkan hasil yang baik bahkan ketika terdapat ketidakseimbangan kelas dalam kumpulan data dengan menggunakan mayoritas tetangga terdekat untuk klasifikasi [7]. Setiap titik data akan dikategorikan sesuai dengan kelas mayoritas dari K tetangga terdekatnya. KNN dipilih karena kesederhanaan implementasi dan kemampuannya beradaptasi terhadap berbagai jenis data.

Berdasarkan konteks ini, penelitian ini menggunakan data dari *platform* media sosial X dan algoritma *K-Nearest Neighbor* untuk menguji opini publik mengenai kinerja Purbaya Yudhi Sadewa dan Sri Mulyani selama bulan pertama mereka menjabat sebagai Menteri Keuangan Indonesia. Studi ini menggunakan metode dan spesifikasi yang seragam untuk kedua figur guna menjamin perbandingan yang adil. Diharapkan penelitian ini akan memberikan ringkasan yang tidak bias mengenai reaksi publik terhadap kedua Menteri Keuangan selama masa

jabatan pertama mereka, membantu memahami bagaimana publik memandang kepemimpinan ekonomi nasional, dan menjadi panduan bagi penelitian terkait di bidang analisis sentimen berdasarkan media sosial.

## 2. Metode Penelitian

Dalam memulai sebuah penelitian, perlu adanya sebuah alur atau tahapan-tahapan yang akan dilakukan. Adapun tahapan-tahapan tersebut ditampilkan dalam bagan alur seperti pada gambar dibawah ini :



Gambar 1. Metodologi Penelitian

### 2.1 Pengumpulan Data

Data yang digunakan bersumber dari *platform* media sosial X dan diperoleh menggunakan bahasa pemrograman *Python* dengan *tools* *Tweet-Harvest*. Kata kunci penelusuran yang digunakan adalah "Purbaya Yudhi Sadewa" dan "Sri Mulyani" pada periode 8 September 2025 hingga 8 Oktober 2025 untuk Purbaya Yudhi Sadewa, serta 21 Oktober 2024 hingga 20 November 2024 untuk Sri Mulyani. Pemilihan periode ini didasarkan pada satu bulan pertama masa jabatan masing-masing sebagai Menteri Keuangan dalam Kabinet Merah Putih, yang merupakan fase krusial untuk mengukur persepsi awal publik terhadap kepemimpinan mereka dalam konteks kabinet yang sama. Proses ini menghasilkan 2960 *tweet* untuk Purbaya Yudhi Sadewa dan 2071 *tweet* untuk Sri Mulyani.

### 2.2 Preprocessing

Tahap *preprocessing* dilakukan untuk meningkatkan kualitas data dengan melakukan pembersihan dan standarisasi teks agar siap dianalisis. Tahap ini sangat penting dalam pengolahan data karena dapat meningkatkan akurasi model dalam menginterpretasikan sentimen pengguna dengan lebih baik [8]. Langkah-langkah *preprocessing* dalam penelitian ini mencakup:

1. *Case Folding*; untuk menstandarisasi format teks agar pemrosesan data lebih sederhana, *case folding* melibatkan pengubahan setiap kata menjadi huruf kecil.
2. *Cleaning*; melibatkan penghapusan tanda baca, karakter khusus, angka, dan *URL* atau tautan yang tidak relevan dengan analisis sentimen.
3. *Tokenizing*; adalah proses membagi setiap kata dalam teks menjadi kata-kata diskrit, atau token, untuk menyederhanakan analisis data.
4. *Stopword Removal*; dalam analisis sentimen, penghapusan kata umum (*stopword removal*) adalah proses menghilangkan kata-kata umum atau kata penghubung yang tidak memiliki banyak makna.
5. *Stemming*; adalah proses penyederhanaan kata-kata dengan makna serupa dengan mengurangi setiap kata dalam teks ke bentuk dasarnya [9].

### 2.3 Pelabelan Data

Tahap pelabelan data dilakukan untuk memberikan label sentimen secara otomatis pada teks yang berbahasa Indonesia. Proses ini menggunakan model *IndoBERT*, yaitu model *transformer* yang telah dilatih secara khusus pada korpus Bahasa Indonesia sehingga mampu memahami konteks dan makna kata dengan baik serta menghasilkan representasi kata yang lebih kontekstual untuk analisis teks berbahasa Indonesia [10].

Dalam penelitian ini, *IndoBERT* digunakan sebagai alat pelabelan otomatis untuk menentukan kecenderungan sentimen pada setiap teks. Secara bawaan, model *IndoBERT* menghasilkan tiga kelas sentimen, yaitu positif, netral, dan negatif. Namun, pada penelitian ini hasil pelabelan disederhanakan menjadi dua kelas, yaitu positif dan negatif. Jika model memprediksi kelas netral, maka label akan disesuaikan berdasarkan probabilitas yang lebih tinggi antara positif dan negatif.

#### 2.4 Pembobotan Kata

Berdasarkan konteks ini, penelitian ini menggunakan data dari *platform* media sosial X dan algoritma *K-Nearest Neighbor* untuk menguji opini publik mengenai kinerja Purbaya Yudhi Sadewa dan Sri Mulyani selama bulan pertama mereka menjabat sebagai Menteri Keuangan Indonesia. Studi ini menggunakan metode dan spesifikasi yang seragam untuk kedua figur guna menjamin perbandingan yang adil. Diharapkan penelitian ini akan memberikan ringkasan yang tidak bias mengenai reaksi publik terhadap kedua Menteri Keuangan selama masa jabatan pertama mereka, membantu memahami bagaimana publik memandang kepemimpinan ekonomi nasional, dan menjadi panduan bagi penelitian terkait di bidang analisis sentimen berdasarkan media sosial. [11].

Pendekatan ini mempertimbangkan dua faktor: *Inverse Document Frequency* (IDF), yang menilai signifikansi suatu kata berdasarkan kelangkaannya di seluruh koleksi dokumen, dan *Term Frequency* (TF), yang mengukur seberapa sering suatu istilah muncul dalam dokumen. Hasil perhitungan TF-IDF berupa vektor numerik dimana setiap kata memiliki bobot yang mencerminkan tingkat kepentingannya. Rumus TF-IDF dapat dilihat pada persamaan berikut:

$$TF - IDF(t, d) = TF(t, d) \times IDF(t) \quad (1)$$

#### 2.5 Klasifikasi KNN

Data dibagi menjadi 20% data pengujian dan 80% data pelatihan. Data pengujian digunakan untuk menilai kinerja klasifikasi, sedangkan data pelatihan digunakan sebagai dasar referensi untuk menemukan tetangga terdekat. Teknik *K-Nearest Neighbor* (KNN) digunakan untuk melakukan kategorisasi. KNN adalah teknik pembelajaran terawasi yang menggunakan kelas mayoritas dari K tetangga terdekatnya dalam data pelatihan untuk mengklasifikasikan data baru. KNN adalah algoritma pembelajaran malas yang tidak memerlukan fase pelatihan, berbeda dengan algoritma lainnya [12]. Vektor fitur dan klasifikasi dari data pelatihan disimpan oleh algoritma ini. Ketika data baru diterima, KNN menentukan K titik terdekat dengan menghitung jarak antara data baru dan semua data pelatihan.

Dalam penelitian ini, klasifikasi KNN diimplementasikan dengan parameter  $K=5$ , yang berarti setiap data uji akan diklasifikasikan berdasarkan kelas mayoritas dari 5 tetangga terdekatnya. Untuk mengukur kedekatan atau kemiripan antar data, digunakan metrik *Cosine Similarity*. *Cosine Similarity* merupakan metrik yang populer dalam analisis teks karena tidak bergantung pada besarnya vektor fitur dan sangat efektif untuk data berbentuk vektor seperti hasil TF-IDF [13]. Metrik ini menghitung tingkat kemiripan antara dua vektor berdasarkan sudut yang dibentuk, dimana nilai *cosine* yang semakin besar mengindikasikan tingkat kemiripan yang semakin tinggi. Tahapan klasifikasi menggunakan KNN adalah sebagai berikut:

1. Menentukan nilai K, yaitu jumlah tetangga terdekat yang akan digunakan untuk klasifikasi.
2. Menghitung jarak atau kemiripan antara data uji dengan seluruh data training menggunakan *Cosine Similarity*. Perhitungan dilakukan menggunakan rumus berikut :

$$Similarity = \cos(\theta) = \frac{\sum_{i=1}^p A_i \times B_i}{\sqrt{\sum_{i=1}^p A_i^2} \times \sqrt{\sum_{i=1}^p B_i^2}} \quad (2)$$

3. Mengurutkan nilai kemiripan dari yang terbesar hingga terkecil, karena nilai *cosine similarity* yang lebih besar menunjukkan tingkat kemiripan yang lebih tinggi.
4. Memilih  $K=5$  data *training* dengan nilai kemiripan tertinggi sebagai tetangga terdekat dari data uji.
5. Menentukan kelas prediksi berdasarkan voting mayoritas dari kelas 5 tetangga terdekat tersebut. Kelas dengan jumlah suara terbanyak akan menjadi hasil prediksi untuk data uji.

#### 2.6 Evaluasi Model

Setelah proses klasifikasi pada data testing selesai dilakukan, hasil prediksi akan dievaluasi menggunakan *Confusion Matrix* untuk mengukur performa model KNN. *Confusion matrix* merupakan tabel yang berfungsi untuk menampilkan jumlah data uji yang berhasil diklasifikasikan dengan tepat maupun tidak tepat, yang berguna untuk memudahkan evaluasi terhadap tingkat akurasi dari suatu sistem klasifikasi [14]. Ilustrasi tabel *Confusion Matrix* dapat dilihat pada Tabel 1.

Tabel 1. *Confusion Matrix*

Kelas Sebenarnya	Kelas Prediksi	
	Negatif	Positif
Negatif	TN	FP
Positif	FN	TP

Dari tabel di atas maka akan dilakukan evaluasi performa model dengan menghitung beberapa metrik evaluasi sebagai berikut:

1. *Accuracy*: merupakan rasio prediksi yang benar terhadap keseluruhan data yang diuji. Metrik ini mengukur seberapa tepat model dalam mengklasifikasikan data secara keseluruhan, baik untuk kelas positif maupun negatif. *Accuracy* dapat dihitung menggunakan rumus berikut:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (3)$$

2. *Precision*: merupakan rasio prediksi benar positif terhadap keseluruhan data yang diprediksi positif. Metrik ini mengukur seberapa akurat model dalam memprediksi kelas positif dari seluruh prediksi positif yang dibuat. *Precision* dapat dihitung menggunakan rumus berikut:

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

3. *Recall*: merupakan rasio prediksi benar positif terhadap keseluruhan data yang sebenarnya positif. Metrik ini mengukur kemampuan model dalam menemukan seluruh data positif yang ada dalam *dataset*. *Recall* dapat dihitung menggunakan rumus berikut:

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

4. *F1-Score*: merupakan rata-rata harmonis dari *precision* dan *recall* yang memberikan keseimbangan antara kedua metrik tersebut. Metrik ini berguna ketika terdapat ketidakseimbangan antara *precision* dan *recall*. *F1-Score* dapat dihitung menggunakan rumus berikut:

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (6)$$

### 3. Hasil dan Pembahasan

#### 3.1 Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan *tweet* berbahasa Indonesia yang dikumpulkan dari *platform* X (*Twitter*) terkait dua tokoh Menteri Keuangan, yaitu Purbaya Yudhi Sadewa dan Sri Mulyani. Pengumpulan data dilakukan pada bulan pertama masa jabatan masing-masing tokoh, yaitu periode 8 September hingga 7 Oktober 2025 untuk Purbaya Yudhi Sadewa, dan periode 21 Oktober hingga 20 November 2024 untuk Sri Mulyani. Total data yang berhasil dikumpulkan sebanyak 5.031 *tweet*, dengan rincian 2.960 *tweet* untuk Purbaya Yudhi Sadewa dan 2.071 *tweet* untuk Sri Mulyani. Data *tweet* yang dikumpulkan dapat dilihat pada Tabel 2 dan Tabel 3.

Tabel 2. Data *Tweet* Purbaya Yudhi Sadewa

No	Tweet
1	Wamen Koperasi Ferry Juliantono Kepala Lembaga Penjamin Simpanan Purbaya Yudhi Sadewa Kepala Badan Penyelenggara (BP) Haji Mochamad Irfan Yusuf mendatangi Istana Kepresidenan Jakarta ditengah isu pelantikan kepala badan dan reshuffle. Mereka tampak mengenakan setelan jas <a href="https://t.co/TtzaoaJ1lr">https://t.co/TtzaoaJ1lr</a>

- 2 Siap lah nggak pernah nggak siap kata Ketua Dewan Komisioner Lembaga Penjamin Simpanan (LPS) Purbaya Yudhi Sadewa saat ditanya apakah siap dilantik di Kompleks Istana Kepresidenan Jakarta Senin (8/9/2025). <https://t.co/h0Xh5P9zsL>
- 3 Purbaya Yudhi Sadewa yang saat ini menjabat Ketua DK Lembaga Penjaminan Simpanan (LPS) tampak hadir di Istana. <https://t.co/KEJXY2DtNW>
- 4 Presiden Prabowo Subianto me-reshuffle sejumlah menteri di Kabinet Merah Putih. Salah satunya Sri Mulyani direshuffle dari jabatannya sebagai Menteri Keuangan (Menkeu). Jabatan Menkeu kini diduduki oleh Purbaya Yudhi Sadewa yang sebelumnya menjabat sebagai Ketua Dewan Komisioner <https://t.co/86csUka04U>
- ... ...
- 2.960 Menkeu Purbaya heran 90% busana muslim di Indonesia produk dari China Dampaknya banyak industri Textile Indonesia bangkrut <https://t.co/G4NQPJtbM> Menteri Keuangan Purbaya Yudhi Sadewa menyinggung banyaknya produk China menguasai pasar Indonesia termasuk produk busana Muslim.

Tabel 3. Data *Tweet* Sri Mulyani

No	Tweet
1	Omongan di pidato dan pilihan menterinya ga sinkron. Bu Retno ga lanjut tapi Sri Mulyani lanjut <a href="https://t.co/M9xs6uUAmM">https://t.co/M9xs6uUAmM</a>
2	Menteri Perempuan Kabinet Prabowo Subianto : Sri Mulyani Menteri Keuangan Lahir: Lampung 26 Agustus 1962 Jabatan Sebelumnya: Petahana Lulusan : Universitas Indonesia University of Illinois Urbana Champaign Kedekatan Politik: Tempo menulis Sri disoronganJoko Widodo dan SBY
3	sri mulyani ril yang punya negara
4	Bagus bener rejeki Sri Mulyani
...	...
2.071	@sweetalani1 Dipikir2 SMI ini menjabat sejak zaman SBY. Lalu jadi menkeu saat JKW Selama itu pula ga terlihat prestasi buat rakyat kecil. Jadi Sri mulyani ini utusan siapa sebenarnya?

### 3.2 Preprocessing

*Preprocessing* dilakukan pada 5.031 data *tweet* melalui lima tahapan pembersihan data, yaitu *case folding*, *cleaning*, *tokenization*, *stopword removal*, dan *stemming*. Contoh hasil *preprocessing* dapat dilihat pada Tabel 4 untuk data Purbaya Yudhi Sadewa dan Tabel 5 untuk data Sri Mulyani.

Tabel 4. Hasil *Preprocessing* Data *Tweet* Purbaya Yudhi Sadewa

Sebelum	Sesudah
Wamen Koperasi Ferry Juliantono	['wamen', 'koperasi', 'ferry', 'juliantono',
Kepala Lembaga Penjamin Simpanan	'kepala', 'lembaga', 'jamin', 'simpan',
Purbaya Yudhi Sadewa Kepala Badan	'purbaya', 'yudhi', 'sadewa', 'kepala',
Penyelenggara (BP) Haji Mochamad	'badan', 'selenggara', 'bp', 'haji',
Irfan Yusuf mendatangi Istana	'mochamad', 'irfan', 'yusuf', 'datang',
Kepresidenan Jakarta ditengah isu	'istana', 'presiden', 'jakarta', 'tengah', 'isu',
pelantikan kepala badan dan reshuffle.	'lantik', 'kepala', 'badan', 'reshuffle',
Mereka tampak mengenakan setelan jas	'tampak', 'kena', 'setel', 'jas']
<a href="https://t.co/TtzaaoJlIr">https://t.co/TtzaaoJlIr</a>	

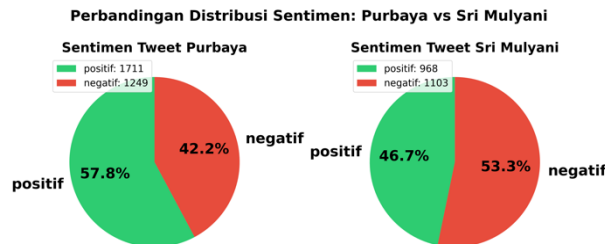
Tabel 5. Hasil *Preprocessing* Data *Tweet* Sri Mulyani



Sebelum	Sesudah
Omongan di pidato dan pilihan menteri ga sinkron. Bu Retno ga lanjut tapi Sri Mulyani lanjut <a href="https://t.co/M9xs6uUAm">https://t.co/M9xs6uUAm</a>	['omong', 'pidato', 'pilih', 'menteri', 'ga', 'sinkron', 'bu', 'retno', 'ga', 'lanjut', 'sri', 'mulyani', 'lanjut']

### 3.3 Pelabelan Data

Pelabelan data dilakukan menggunakan model *IndoBERT* untuk mengklasifikasikan sentimen setiap *tweet* secara otomatis menjadi dua kategori, yaitu positif dan negatif. Distribusi sentimen untuk masing-masing tokoh dapat dilihat pada Gambar 2.



Gambar 2. Perbandingan Distribusi Sentimen: Purbaya vs Sri Mulyani

Berdasarkan Gambar 2, terlihat perbedaan distribusi sentimen antara kedua tokoh. *Tweet* yang membahas Purbaya Yudhi Sadewa didominasi oleh sentimen positif dengan persentase 57,8%, sedangkan sentimen negatif sebesar 42,2%. Sementara itu, *tweet* yang membahas Sri Mulyani menunjukkan dominasi sentimen negatif dengan persentase 53,3%, sedangkan sentimen positif sebesar 46,7%. Perbedaan distribusi sentimen ini menunjukkan bahwa respons publik terhadap kedua tokoh berbeda pada bulan pertama masa jabatan mereka sebagai Menteri Keuangan.

### 3.4 Pembobotan Kata

Proses pembobotan kata dilakukan menggunakan metode TF-IDF. Untuk data Purbaya Yudhi Sadewa, terdapat 2.960 dokumen *tweet* yang diubah menjadi matriks TF-IDF, sedangkan untuk data Sri Mulyani terdapat 2.071 dokumen *tweet*. Contoh hasil pembobotan dapat dilihat pada Tabel 6 dan Tabel 7.

Tabel 6. Hasil Pembobotan TF-IDF Tweet Purbaya Yudhi Sadewa

No	Term	TF			IDF	TF-IDF		
		D1	D2	D3		D1	D2	D3
1	badan	2	0	0	1.693	0.339	0	0
2	ketua	0	1	1	1.288	0	0.161	0.266
3	lembaga	1	1	1	1	0.100	0.125	0.207
...	...	...	...	...	...	...	...	...
4.397	wamen	1	0	0	1.693	0.170	0	0

Tabel 7. Hasil Pembobotan TF-IDF Tweet Sri Mulyani

No	Term	TF			IDF	TF-IDF		
		D1	D2	D3		D1	D2	D3
1	agustus	0	1	0	1.693	0	0.182	0
2	lanjut	2	0	0	1.693	0.512	0	0
3	menteri	1	2	0	1.288	0.277	0.125	0
...	...	...	...	...	...	...	...	...
4.930	universitas	0	1	0	1.693	0	0.182	0

### 3.5 Klasifikasi KNN

Sebelum membangun model klasifikasi, data dibagi menjadi 80% data *training* dan 20% data *testing* untuk masing-masing tokoh. Pembagian data dilakukan secara proporsional untuk mempertahankan distribusi kelas pada setiap subset. Rincian pembagian data dapat dilihat pada Tabel 8.

Tabel 8. Pembagian Data untuk Klasifikasi KNN

Tokoh	Total Data	Data Training (80%)	Data Testing (20%)
Purbaya Yudhi Sadewa	2.960	2.368	592
Sri Mulyani	2.071	1.657	414
<b>Total</b>	<b>5.031</b>	<b>4.025</b>	<b>1.006</b>

Setelah data dibagi, klasifikasi dilakukan menggunakan algoritma *K-Nearest Neighbor* (KNN) dengan parameter  $K=5$  dan metrik *cosine similarity*. Proses klasifikasi dilakukan pada 592 data *testing* untuk Purbaya Yudhi Sadewa dan 414 data *testing* untuk Sri Mulyani.

Sebagai ilustrasi, berikut contoh proses klasifikasi untuk satu data *testing* dari dataset Purbaya Yudhi Sadewa dengan indeks 999. Setelah melalui tahap *preprocessing*, data testing memiliki teks: "*menkeu purbaya yudhi sadewa punya hobi koleksi keris salah satu keris mar keris mar mesem*" dengan label sebenarnya adalah positif.

Sistem kemudian menghitung *cosine similarity* berdasarkan vektor TF-IDF antara data *testing* dengan seluruh data *training* untuk mencari 5 tetangga terdekat. Hasil perhitungan *similarity* dapat dilihat pada Tabel 9.

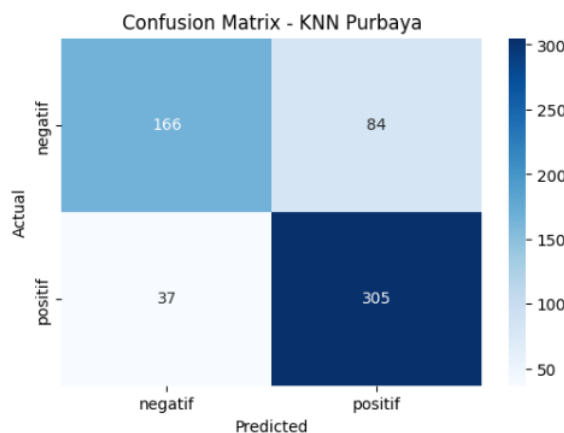
Tabel 9. Lima Tetangga Terdekat untuk Data Indeks 1657

Indeks Data	Cosine Similarity	Label Sentimen
432	0.601	Positif
1611	0.206	Negatif
285	0.123	Positif
1332	0.113	Positif
985	0.104	Positif

Berdasarkan hasil voting dari 5 tetangga terdekat pada Tabel 9, terdapat 4 label positif dan 1 label negatif, sehingga data testing dengan indeks 999 diklasifikasikan sebagai sentimen positif berdasarkan voting mayoritas. Hasil prediksi ini sesuai dengan label sebenarnya, yang menunjukkan bahwa model KNN berhasil mengklasifikasikan data dengan tepat pada kasus ini.

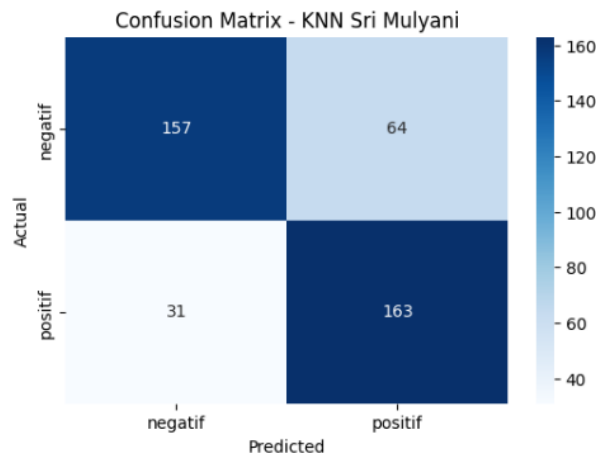
### 3.6 Evaluasi Model

Proses evaluasi dilakukan untuk mengukur performa klasifikasi KNN pada kedua *dataset* menggunakan data *testing*. Evaluasi dilakukan dengan *confusion matrix* dan *classification report* yang menampilkan metrik akurasi, *precision*, *recall*, dan *f1-score*.



Gambar 3. *Confusion Matrix* Model KNN Purbaya Yudhi Sadewa





Gambar 4. *Confusion Matrix* Model KNN Sri Mulyani

Gambar 3 dan Gambar 4 menampilkan *confusion matrix* untuk masing-masing model. *Confusion matrix* menunjukkan distribusi hasil klasifikasi antara prediksi dan label sebenarnya, yang terdiri dari *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). Berdasarkan *confusion matrix* tersebut, dapat dihitung nilai *precision*, *recall*, *f1-score*, dan *accuracy* yang ditampilkan dalam *classification report*.

Classification Report Purbaya:			
	precision	recall	f1-score
negatif	0.82	0.66	0.73
positif	0.78	0.89	0.83
accuracy			0.80
macro avg	0.80	0.78	0.78
weighted avg	0.80	0.80	0.79

Gambar 5. *Classification Report* Model KNN Purbaya Yudhi Sadewa

Classification Report Sri Mulyani:			
	precision	recall	f1-score
negatif	0.84	0.71	0.77
positif	0.72	0.84	0.77
accuracy			0.77
macro avg	0.78	0.78	0.77
weighted avg	0.78	0.77	0.77

Gambar 6. *Classification Report* Model KNN Sri Mulyani

Berdasarkan *classification report* pada Gambar 5 dan Gambar 6, performa kedua model dapat dibandingkan secara ringkas pada Tabel 10.

Tabel 10. Perbandingan Hasil *Classification Report* Kedua Tokoh

Tokoh	Accuracy	Precision	Recall	F1-Score
Purbaya Yudhi Sadewa	80%	80%	78%	79%
Sri Mulyani	77%	77%	77%	77%

#### 4. Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa analisis sentimen publik terhadap kinerja satu bulan pertama Purbaya Yudhi Sadewa dan Sri Mulyani sebagai Menteri Keuangan Indonesia pada data media sosial X menunjukkan distribusi sentimen yang berbeda. Purbaya Yudhi Sadewa mendapatkan sentimen positif sebesar 57,8% dan negatif 42,2% dari 2.960 *tweet*, sementara Sri Mulyani mendapatkan sentimen positif

sebesar 46,7% dan negatif 53,3% dari 2.071 *tweet*. Perbedaan distribusi ini mengindikasikan adanya variasi respons publik terhadap kedua tokoh pada periode awal masa jabatan mereka.

Model *K-Nearest Neighbor* yang diimplementasikan dengan parameter  $K=5$  dan metrik *cosine similarity* menunjukkan performa yang baik dalam klasifikasi sentimen. Model KNN untuk Purbaya Yudhi Sadewa mencapai *accuracy* sebesar 80% dengan *precision* 80%, *recall* 78%, dan *f1-score* 79%, sedangkan model untuk Sri Mulyani mencapai *accuracy* 77% dengan *precision* 78%, *recall* 77%, dan *f1-score* 77%. Hasil ini mengindikasikan bahwa model memiliki kemampuan yang baik dalam mengklasifikasikan sentimen pada kedua *dataset* dengan performa yang seimbang antar metrik evaluasi.

## Daftar Rujukan

- [1] K. Y. Rahayu, "Purbaya Yudhi Sadewa Dilantik Jadi Menkeu, Gantikan Sri Mulyani," *Kompas.id*, Jakarta, Sep. 08, 2025. Accessed: Nov. 01, 2025. [Online]. Available: <https://www.kompas.id/artikel/purbaya-yudhi-sadewa-dilantik-jadi-menkeu-gantikan-sri-mulyani>
- [2] D. S. Rahmah and Y. Abbas, "Political turnover and business decision: A study of Indonesian public companies," *Journal of Accounting and Investment*, vol. 24, no. 3, pp. 634–653, Jun. 2023, doi: 10.18196/jai.v24i3.17565.
- [3] M. Riski Andika Rambe, I. Zufria, M. Ikhsan Rifki, P. Studi Ilmu Komputer, and F. Sains dan Teknologi, "Analisis Sentimen Masyarakat pada Platform Media Sosial X (Twitter) terhadap Pelantikan Kabinet Merah Putih Menggunakan Bernoulli Naïve Bayes," vol. 6, no. 1, 2025.
- [4] R. Y. Sari, N. Diyah and Mardhiyatna, "Analisis Sentimen Digital Publik terhadap Perguruan Tinggi Negeri Menggunakan Google Review: Studi Kasus ITERA," *Jurnal Pustaka AI*, vol. 5, no. 2, pp. 298-305, 2025.
- [5] L. O. Lukmana, "ANALISIS SENTIMEN TERHADAP ULASAN PENGGUNA APLIKASI THREADS INSTAGRAM DI PLAYSTORE MENGGUNAKAN ALGORITMA NAIVE BAYES," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 13, no. 2, Apr. 2025, doi: 10.23960/jitet.v13i2.6250.
- [6] P. Sayarizki and H. Nurrahmi, "Implementation of IndoBERT for Sentiment Analysis of Indonesian Presidential Candidates," *Journal on Computing*, vol. 9, no. 2, pp. 61–72, 2024, doi: 10.34818/indojc.2024.9.2.934.
- [7] A. Rhamadanti, A. Rifa'i, F. Dikananda, and K. Anam, "ANALISIS SENTIMEN PADA ULASAN ACCESS BY KERETA API INDONESIA DENGAN K-NEAREST NEIGHBOR," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 1, pp. 2830–7062, 2024, doi: 10.23960/jitet.v12i1.3691.
- [8] I. G. S. D. Putra and I. N. T. A. Putra, "IMPLEMENTASI METODE NAÏVE BAYES PADA ANALISIS SENTIMEN PENGGUNA APLIKASI MOBILE KITA BISA," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 13, no. 2, Apr. 2025, doi: 10.23960/jitet.v13i2.6423.
- [9] A. D. Prastiko and A. D. Wiranata, "Analisis Sentimen Publik terhadap Fenomena Judi Online di Media Sosial X dengan SVM," *Pustaka AI*, vol. 5, no. 2, pp. 306-315, 2025.
- [10] J. Pranata, S. Agustian, J. Jasril, and E. Haerani, "Penggunaan Model Bahasa indoBERT pada metode Random Forest untuk Klasifikasi Sentimen dengan Dataset Terbatas," *Building of Informatics, Technology and Science (BITS)*, vol. 6, no. 3, pp. 1668–1676, Dec. 2024, doi: 10.47065/bits.v6i3.6335.
- [11] D. Septiani and I. Isabela, "SINTESIA: Jurnal Sistem dan Teknologi Informasi Indonesia ANALISIS TERM FREQUENCY INVERSE DOCUMENT FREQUENCY (TF-IDF) DALAM TEMU KEMBALI INFORMASI PADA DOKUMEN TEKS," 2023.
- [12] R. Bagus Trianto *et al.*, "Kombinasi Metode K-Nearest Neighbor dengan Cosine Similarity untuk Prediksi Serangan Firewall pada Jaringan Komputer," vol. 6, no. 4, pp. 672–679, 2021, doi: 10.32493/informatika.v6i4.12680.
- [13] N. L. Putri, B. Warsito, and B. Surarso, "Pengaruh Klasifikasi Sentimen Pada Ulasan Produk Amazon Berbasis Rekayasa Fitur dan K-Nearest Neighbor," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 11, no. 1, pp. 65–74, Feb. 2024, doi: 10.25126/jtiik.20241117376.
- [14] R. Nurhidayat and K. E. Dewi, "KOMPUTA : Jurnal Ilmiah Komputer dan Informatika PENERAPAN ALGORITMA K-NEAREST NEIGHBOR DAN FITUR EKSTRAKSI N-GRAM DALAM ANALISIS SENTIMEN BERBASIS ASPEK," vol. 12, no. 1, 2023, [Online]. Available: <https://www.kaggle.com/datasets/hafidahmusthaanah/skincare-review?select=00.+Review.csv>.

-----