

## ANALISIS KLASIFIKASI SENTIMEN PENGGUNA MEDIA SOSIAL TWITTER TERHADAP PENUNDAAN PEMILU PRESIDEN TAHUN 2024

Jimmy <sup>1\*</sup>, Eni Heni Hermaliani <sup>2</sup>, Laela Kurniawati <sup>3</sup>

<sup>1\*</sup> Program Studi Ilmu Komputer, Fakultas Teknologi Informasi, Universitas Nusa Mandiri, Kota Jakarta Pusat, Daerah Khusus Ibukota Jakarta, Indonesia.

<sup>2,3</sup> Program Studi Sistem Informasi, Fakultas Teknologi Informasi, Universitas Nusa Mandiri, Kota Jakarta Pusat, Daerah Khusus Ibukota Jakarta, Indonesia.

Email: 14210178@nusamandiri.ac.id <sup>1\*</sup>, enie\_h@nusamandiri.ac.id <sup>2</sup>, laela@nusamandiri.ac.id <sup>3</sup>

### Histori Artikel:

*Dikirim* 29 Maret 2023; *Diterima dalam bentuk revisi* 15 April 2023; *Diterima* 25 April 2023; *Diterbitkan* 10 Mei 2023. Semua hak dilindungi oleh Lembaga Penelitian dan Pengabdian Masyarakat (LPPM) STMIK Indonesia Banda Aceh.

### Abstrak

Wacana penundaan pemilu 2024 dan perpanjangan masa jabatan presiden mendorong munculnya sentimen di sebagian daerah di Indonesia. Wacana tersebut berimplikasi pada perpanjangan masa jabatan presiden dinilai inkonstitusional. Isu penundaan pemilu ini sudah jelas bukan aspirasi rakyat, melainkan hanya kepentingan nafsu dan syahwat politik di kalangan penguasa yang ingin melanggengkan kekuasaan sehingga dinilai melanggar dan melecehkan konstitusi yang telah mengatur bahwa Pemilu digelar setiap lima tahun sekali. Penelitian ini bertujuan untuk mengetahui bagaimana analisis sentimen penundaan pemilu 2024 terhadap komentar masyarakat Indonesia yang ada di Twitter atau disebut dengan tweet. Jumlah komentar yang akan digunakan pada penelitian sebanyak 1826 yang terdiri dari 710 sentimen positif dan 1116 sentimen negatif. Metode penelitian yang digunakan adalah metode Cross-Industry Standard Process for Data Mining (CRISP-DM). Dengan metode ini, tahapan penelitian yang dilakukan adalah pengumpulan data, pemahaman bisnis, pemahaman data, data preprocessing, pemberian label data, pemodelan, evaluasi, dan deployment. Berdasarkan hasil yang diperoleh pada tahapan modelling dari data tweet yang sudah terkumpul kemudian diolah dan dianalisis sentimennya dengan model train/test split data dan k-fold cross validation menggunakan algoritma Naive Bayes (NB), Support Vector Machine (SVM), Deep Learning (DL), dan Decision Tree (DT) dan perbandingan 80:20 untuk data training dan testing maka nilai akurasi tertinggi didapat dengan model train/test split data menggunakan algoritma Deep Learning menghasilkan akurasi sebesar 80,55%.

**Kata Kunci:** Klasifikasi; Data Mining; Rapidminer; Analisis Sentiment; Twitter.

### Abstract

The discourse of postponing the 2024 election and extending the term of office of the president has stimulated sentiment in some regions in Indonesia. The discourse has implications for the extension of the president's term of office which is considered unconstitutional. The issue of postponing the election is clearly not the aspiration of the people, but only the interests of political passion and lust among the rulers who want to perpetuate power so that it is considered to violate and insult the constitution which stipulates that elections are held every five years. This study aims to determine how to analyze the sentiment of postponement the 2024 election on comments from the Indonesian people on Twitter or called tweets. The number of comments that will be used in the study is 1826 consisting of 710 positive sentiments and 1116 negative sentiments. The research method used is the Cross-Industry Standard Process for Data Mining (CRISP-DM) method. With this method, the stages of research carried out are data collection, business understanding, data understanding, data preprocessing, data labeling, modeling, evaluation, and deployment. Based on the results obtained at the modeling stage, the tweet data that has been collected is then processed and analyzed for sentiment with a train/test split data model and k-fold cross validation using the Naive Bayes (NB) algorithm, Support Vector Machine (SVM), Deep Learning (DL), and Decision Tree (DT) and a comparison of 80:20 for training and testing data, the highest accuracy value is obtained by using the train/test split data model using the Naive Bayes algorithm to produce an accuracy of 80,55%.

**Keyword:** Classification; Data Mining; Rapidminer; Sentiment Analysis; Twitter.

## 1. Pendahuluan

Polemik terkait dengan penundaan Pemilihan Umum 2024 sangat ramai dibicarakan oleh masyarakat Indonesia. Sebagian besar masyarakat menolak wacana tersebut karena melanggar konstitusi, sedangkan sebagian di antaranya mendukung dikarenakan adanya program pembangunan yang belum selesai pasca covid-19. Prof. Nurliah Nurdin, M.A. yang merupakan salah satu narasumber mengatakan bahwa pemilu 2024 harus tetap berjalan. Namun demikian, faktanya adalah bahwa posisi pemerintah (eksekutif) saat ini berada dalam *majority*. Kondisi ini disebabkan oleh dukungan koalisi partai di legislatif (parlemen) yang kuat dalam sistem *presidential*. Hal ini disebut dengan *unified government*. *Unified government* akan memuluskan setiap wacana yang ada untuk diangkat menjadi sesuatu yang nyata. Contoh lain dari hal ini adalah pelaksanaan amandemen konstitusi, sebagai kesepakatan lembaga eksekutif dan legislatif [1].

Ketua PP Muhammadiyah Bidang Hukum dan HAM, Busyro Muqoddas menyebut para elite politik yang mengusulkan penundaan Pemilu 2024 tak ubahnya seperti keledai yang tak punya rasa malu. Isu penundaan Pemilu yang digulirkan ini menunjukkan semakin vulgarnya sikap penguasa yang tidak punya rasa malu, mereka seperti keledai-keledai politik yang tidak belajar dari masa lalu. Ide penundaan Pemilu 2024 yang berimplikasi pada perpanjangan masa jabatan presiden dinilai inkonstitusional. Di samping itu, tidak ada satu alasan pun yang bisa dibenarkan untuk mendukung ide ini. Isu penundaan pemilu ini sudah jelas bukan aspirasi rakyat. Melainkan hanya kepentingan nafsu dan syahwat politik di kalangan penguasa yang ingin melanggengkan kekuasaan. Kita tidak boleh membiarkan ini terjadi. Sepatunya elemen-elemen masyarakat sipil, ormas, NGO, hingga organisasi kemahasiswaan dengan tegas menyatakan penolakan terhadap ide penundaan pemilu ini [2].

Guru Besar Hukum Tata Negara, Denny Indrayana, mengkritik keras partai-partai pendukung pemerintah yang secara terbuka mengusulkan Penundaan Pemilu, berarti pula perpanjangan jabatan presiden dan parlemen, serta kepala daerah. Usul tersebut dinilainya melanggar dan melecehkan konstitusi yang telah mengatur bahwa Pemilu digelar setiap lima tahun sekali. Ini nyata-nyata adalah potret pelanggaran konstitusi berjamaah yang didasari pada dahaga atas kekuasaan semata. Dalam teori ketatanegaraan, pengecualian tidak mengikuti aturan konstitusi itu hanya dimungkinkan dalam situasi sangat darurat, dan hanya demi menyelamatkan negara dari ancaman serius yang berpotensi menghilangkan negara. Sejarah Indonesia mencatat, pembubaran Konstituante dan kembali ke UUD 1945, melalui Dekrit Presiden 5 Juli 1959, sebagai salah satu pelanggaran konstitusi, yang akhirnya diakui menjadi sumber hukum bernegara yang sah dan berlaku. Namun, alasan pelanggaran konstitusi harus jelas untuk penyelamatan negara dan melindungi seluruh rakyat Indonesia Saat ini, Indonesia tidak sedang mengalami kondisi darurat [3].

Pakar hukum tata negara dari Universitas Gadjah Mada Zainal Arifin Mochtar mengkhawatirkan dua skenario yang bisa saja dibuat untuk meloloskan wacana penundaan Pemilu 2024. Pertama, Zainal khawatir akan kemungkinan agenda amandemen UUD 1945 yang sedang dibahas MPR disusupi agenda perpanjangan masa jabatan presiden. Skenario kedua, Zainal Arifin Mochtar khawatir ada upaya menciptakan kondisi seakan-akan obyektif, rasional, dan konstitusional untuk melakukan amandemen dan mengubah masa jabatan presiden dengan memperalat KPU [4]. Kekhawatiran adanya upaya serius menunda pemilu dan memperpanjang masa jabatan presiden juga disampaikan pakar hukum lainnya, diantaranya Denny Indrayana, Yusril Selanjutnya, penundaan Pilkada harus menerbitkan Peraturan Pemerintah Pengganti Undang-Undang (PERPPU) sebagai dasar hukum dari adanya penundaan Pilkada. Pemilu paralel 2024 yang menggabungkan tahapan pemilu serentak pada sejumlah Provinsi atau Kabupaten Kota bisa menyebabkan beban kerja pada penyelenggara dan kelelahan prosedural dalam pemilu yang akan berdampak kepada laporan kode etik oleh para penyelenggara Pemilu 2024 mendatang [5]. Sehingga, hal ini menimbulkan kepada pentingnya memisahkan pemilu lokal dan pemilu nasional dalam agenda pemilu serentak yang mana perlu dipertimbangkan kembali bagi penyelenggara pemilu pasca Pemilu 2024 yang akan datang. Konsep Pemilu serentak, diperlukan sebuah kebijakan yang dilakukan untuk

dapat mengatur ulang, melakukan sinkronisasi dan kodifikasi UU Pemilu, termasuk Pilkada yang mana harus adanya materi teknis dan pengelolaan Pilkada dengan penjabaran pemilu (termasuk Pilkada), latar belakang penyelenggara pilkada, jadwal pelaksanaan Pilkada waktu pemungutan suara, metode penetapan paslon terpilih, dan jadwal pelantikan Kepala Daerah beserta Wakil terpilih. Pemilu serentak yang dilaksanakan sebelumnya, pada tahun 2019 terdapat masalah yang menggemparkan, salah satunya yaitu penyelenggara pemilu menjadi korban jiwa sebagai akibat dari penyelenggaraan pemilu serentak serta problem teknis lainnya.

Perkembangan penggunaan Twitter sangat cepat. Hal ini berdasarkan menurut laporan statista dari sumber yang dilansir situs databoks.katadata.co.id [6] pada 23 Maret 2022, yaitu terdapat lebih dari 18,45 juta pengguna aktif di seluruh dunia bersama-sama mengirimkan 250 juta tweet setiap hari. Pengguna Twitter berasal dari Indonesia hingga Maret 2022 menjadi urutan kelima yang terbanyak di dunia. Kesederhanaan dan kemudahan dalam penggunaan merupakan beberapa alasan mengapa Twitter lebih digemari masyarakat Indonesia dalam berkomunikasi. Setiap pengguna Twitter bebas mem-*post* tweet dengan batasan 140 karakter.

Tweet adalah teks status pengguna yang digunakan untuk memberikan informasi di Twitter. Berdasarkan kutipan hasil penelitian [7], *tweets* yang merupakan hasil dari penyaluran opini dan komentar, merupakan resource yang dapat digunakan untuk menganalisis sentimen khalayak umum terhadap suatu instansi maupun perorangan. Hal tersebut disebabkan karena tweets tersebut mengandung sentimen yang dapat dijadikan sebagai tolak ukur pandangan khalayak umum yang dapat dijadikan sebagai bahan evaluasi ke depannya. Dalam menentukan sentimen suatu *tweet*, dapat dilakukan dengan mengklasifikasikannya ke dalam tiga kelas, yaitu positif, negatif, dan netral. Opini melalui tweet inilah yang dapat dimanfaatkan untuk melihat bagaimana sentimen yang dimunculkan salah satunya adalah mengenai penundaan pemilu 2024.

Penentuan polaritas positif atau negatifnya suatu komentar dapat dilakukan secara manual, tetapi seiring bertambahnya sumber komentar menjadi semakin banyak tentunya waktu dan usaha yang dibutuhkan untuk mengklasifikasikan polaritas komentar tersebut akan semakin banyak terpakai. Oleh karena itu, diajukan penerapan metode pembelajaran mesin untuk mengklasifikasi polaritas komentar dari sumber data yang sangat banyak tersebut. Untuk melakukan hal itu, bisa menggunakan salah satu fungsi dari *text mining*, dalam hal ini adalah klasifikasi dokumen. Beberapa peneliti sebelumnya telah melakukan penelitian untuk mengukur sentimen analisis dengan berbagai algoritma, antara lain algoritma *Naive Bayes* (NB) [8], *Support Vector Machine* (SVM) [9], *K-Nearest Neighbour* (K-NN) [10], *Decision Tree* [11], *K-Medoid* [12], dan *Backpropagation Neural Network* (BNN) [13]. Dengan adanya perbedaan beberapa penelitian sebelumnya, mendorong penulis untuk melakukan penelitian dengan membandingkan beberapa metode klasifikasi dengan kasus yang berbeda. Di dalam penelitian ini, akan dibahas tahapan yang dilalui untuk melakukan proses analisis sentimen terhadap komentar tentang penundaan pemilu 2024 pada media sosial Twitter. Dimulai dari tahap *preprocessing* sampai tahap analisis sentimen dengan *Naive Bayes* (NB)[8], *Support Vector Machine* (SVM)[9], *Deep Learning*[14], dan *Decision Tree*[11].

Penelitian oleh menggunakan machine learning untuk membangun sistem pendeteksi rumor Indonesia di Twitter menggunakan klasifikasi SVM dan pemilihan fitur dengan pembobotan TF IDF. Dataset yang digunakan adalah metode crawl dan proses manual labelling. Subjek data yang digunakan adalah tren sekitar dengan 47.499 dataset. Langkah selanjutnya adalah pemilihan fitur, pemodelan N-Gram, dan evaluasi matriks konfusi. Berdasarkan hasil investigasi, sistem bekerja dengan baik dengan nilai akurasi 78,71%[15]. Menggunakan *emoticon* untuk membangun corpus berbahasa Inggris dari *Twitter* dengan sentimen positif, negatif dan netral. Untuk kelas netral Pak dan Paurobek mengambil data training dari tweet akun media berbahasa Inggris. Metode yang digunakan adalah *Naive Bayes* dengan *N-gram*. Performasi terbaik dihasilkan ketika menggunakan *bigram*.

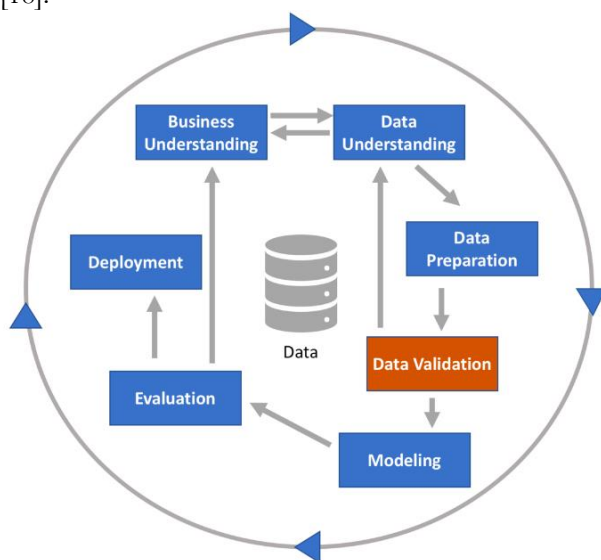
*Twitter* dapat digunakan sebagai platform bagi kandidat dan pengguna untuk mendapatkan jangkauan substansial dalam menunjukkan pandangan mereka kepada dunia. Oleh karena itu, sangat penting untuk menganalisis dan memahami peran yang dapat dimainkan *Twitter* dalam mengukur sentiment seputar isu-isu penting yang digunakan untuk mengeluarkan pendapat dan komentarnya, termasuk komentar tentang calon Presiden Indonesia 2014 [16]. Penelitian ini menampung pendapat

dan komentar masyarakat tersebut dengan membaginya menjadi lima *class attribute*, yaitu sangat positif, positif, netral, negatif dan sangat negatif. Proses klasifikasi pada penelitian ini menggunakan metode klasifikasi *Naïve Bayes Classifier* (NBC) dengan *preprocessing* data menggunakan *tokenisasi*, *cleansing* dan *filtering*. Data yang digunakan adalah *tweet* dalam bahasa Indonesia tentang calon Presiden Indonesia 2014, dengan jumlah dataset sebanyak 900 *tweet* yang didistribusikan secara merata kedalam lima *class attribute*. Hasil akurasi tertinggi didapatkan saat menggunakan *tokenisasi n-gram*, *stopword list* WEKA dan *emoticons*, yaitu dengan nilai akurasi 71,9%, nilai presisi 71,6%, nilai recall 71,9% nilai TP rate 66,1% dan nilai TN rate 65%.

Penelitian menggunakan *machine learning* untuk mengklasifikasikan *movie reviews* [17]. Penelitian ini melakukan klasifikasi sentimen terhadap *review* film dan menentukan apakah *review* film tersebut memiliki sentimen positif ataukah negatif. Fitur yang berbeda dari *review* tersebut diekstrak dan digunakan algoritma *machine learning Naïve Bayes* dan *Support Vector Machine* (SVM) untuk menghasilkan model klasifikasi. Mereka memperoleh akurasi antara 78,7% ketika menggunakan *Naïve Bayes* pada penggunaan unigram. Akurasi yang diperoleh ketika menggunakan SVM dengan unigram adalah 72,8%.

## 2. Metode Penelitian

Metode pada penelitian ini menggunakan model *Cross-Industry Standard Process for Data Mining* (CRISP-DM). CRISP-DM adalah metode yang menyediakan standar baku pada data mining dan dapat diterapkan pada strategi pemecahan masalah umum. CRISP-DM juga merupakan standar proses data mining sebagai strategi pemecahan masalah umum dari bisnis ataupun penelitian. Model CRISP-DM terdiri dari enam fase yaitu *business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation* dan *deployment*[18].



Gambar 1. Diagram Sistem Klasifikasi untuk Analisis Sentimen

Pada penelitian ini terdapat beberapa tahapan seperti yang ditunjukkan pada Gambar 1, penelitian ini menggunakan metode *Cross-Industry Standard Process for Data Mining* (CRISP-DM), dimana data akan diproses dengan melakukan beberapa tahapan seperti berikut :

- 1) Pemahaman Bisnis (*Business Understanding*)

Pada tahapan ini dilakukan penentuan permasalahan yang akan diangkat yaitu analisis sentimen terhadap komentar mengenai penundaan pemilu 2024, langkah yang dilakukan yaitu mencari data komentar yang terdapat pada twitter.

## 2) Pemahaman Data (*Data Understanding*)

Pada tahapan ini peneliti mulai dengan mengumpulkan data awal lalu digabungkan menjadi satu menjadi dataset, data yang telah dikumpulkan menjadi satu akan diteliti dan hasil dari kegiatan pengambilan data dalam rangka untuk mengidentifikasi masalah data yang telah terkumpul. Tahap pengumpulan data ini menggunakan stream twitter API yang telah disediakan oleh twitter dan menggunakan tool rapidminer untuk crawling data. API keys dan access token sangat diperlukan untuk pengambilan data atau stream twitter API, untuk autentifikasi mendaftarkan terlebih dahulu akun twitter ke developer twitter. bertujuan untuk mengumpulkan, mengidentifikasi, dan memahami data yang dimiliki. Data tersebut juga harus dapat diverifikasi kebenarannya. Data yang digunakan dalam penelitian ini adalah data komentar yang terdapat pada twitter mengenai penundaan pemilu 2024,

## 3) Koleksi Data (*Data Collection*)

Pada tahapan ini peneliti melakukan pengambilan data komentar pada twitter pada April 2022. Sebelum melakukan pengambilan data peneliti harus mengkonfigurasi pada tool rapidminer agar pengambilan data bisa dilakukan, peneliti harus memilih data mana saja yang akan diambil seperti rating, tanggal, nama, dan yang terpenting komentar. Proses pengambilan data komentar yang dilakukan pada penelitian ini diambil dengan menggunakan tool rapidminer. Data yang diambil pada penelitian ini hanyalah data komentar mengenai penundaan pemilu 2024.

## 4) Persiapan Data (*Data Preparation*)

Pada tahapan ini data yang sudah didapatkan sebelumnya, akan diberikan sentimen. Setelah data yang telah didapatkan diberikan sentimen, setelah itu data akan masuk tahap *data preprocessing*. Berikut adalah tahapan-tahapan yang dilakukan pada tahapan *data preprocessing*:

- Remove duplicates* merupakan proses untuk menghapus data yang sama, atau diambil secara berulang pada saat kita crawling data.
- Case Folding* merupakan proses untuk mengubah semua data menjadi huruf kecil.
- Tokenizing* merupakan proses pemecahan sekumpulan karakter dalam suatu teks kedalam satuan kata, pada proses ini juga dapat menghilangkan karakter pembatas, menghapus angka, dan menghapus tanda baca atau karakter yang bukan merupakan huruf.
- Stopwords* merupakan proses menghilangkan kata yang tidak memiliki makna atau tidak penting dengan menggunakan stopword list dari Tala. Setelah menggunakan Stopword list tersebut, lalu ditambahkan lagi beberapa kata pada data kita seperti “kok, lah, dll” yang tidak memiliki makna pada penelitian ini [19].
- Term Weighting (TF-IDF)* merupakan proses untuk memberikan pembobotan kata dengan mencari nilai *Term Frequency (TF)*, kemudian mencari nilai *Document Frequency (DF)*, lalu mencari nilai *Invers Document Frequency (IDF)* setelah itu baru menghitung bobot[20].

## 5) Pemodelan (*Modeling*)

Pada tahapan ini akan dilakukan pemodelan terhadap data komentar yang sudah dilakukan *preprocessing*. Data komentar tersebut akan dipecah menjadi data training dan data testing, pada tahapan ini akan dipecah menjadi dua tahapan yaitu *split data* dan *cross fold validation*. Pada *split data* peneliti akan membagi data komentar tersebut dengan perbandingan *data training* dan *data testing* yaitu 80:20. Lalu pada tahap *cross fold validation* kita akan membagi secara acak kedalam 10 bagian. Kemudian pada tahap pemodelan ini peneliti menggunakan algoritma *naive bayes*, *support vektor machine*, *deep learning* dan *decision tree* untuk mendapatkan nilai sentimen positif dan sentimen negatif dari sebuah data komentar.

## 6) Evaluasi (*Evaluation*)

Pada tahapan ini akan dilakukan evaluasi metode klasifikasi dengan mengukur performa menggunakan *confusion matrix* terhadap algoritma *naive bayes*, *support vektor machine*, *deep learning* dan *decision tree*.

## 7) Kesimpulan dan Saran

Pada tahapan ini akan dilakukan pemberian kesimpulan yang telah didapat selama melakukan penelitian ini serta memberikan saran untuk penelitian lanjutan yang akan dilakukan oleh peneliti lain.



### 3. Hasil dan Pembahasan

Pada bab ini akan dijelaskan mengenai hasil dan pembahasan dari performa dari masing-masing algoritma dengan metode yang digunakan untuk penelitian ini yaitu CRISP-DM. Penelitian ini menggunakan data yang diambil melalui bantuan Twitter Data API. Setelah melewati proses seleksi didapatkan 1.826 data komentar. Selanjutnya dilakukan pelabelan dengan label positif dan label negatif secara manual.

#### 1) Business Understanding

Tahap ini merupakan tahap pemahaman mengenai objek penelitian yang dilakukan. Penelitian dilakukan dengan mencari informasi melalui media Twitter, kemudian mencari komentar pada komentar mengenai penundaan pemilu 2024. Media Twitter tidak hanya sebagai media informasi, akan tetapi dapat digunakan untuk melihat sentimen mengenai sebuah komentar yang dimuat setiap akun status.

#### 2) Data Understanding

Pada tahap ini, peneliti melakukan proses pengambilan data diperoleh melalui *Twitter*. Dataset penelitian berupa teks komentar *Twitter* bahasa Indonesia mengenai penundaan pemilu 2024. Total komentar yang didapatkan dari akun status yang dimuat dari bulan April hingga Mei 2022 sebanyak 3035 komentar. kemudian pada tahap ini peneliti juga melakukan penentuan keywords yang akan digunakan untuk proses crawling atau pengumpulan data yang saling berkaitan tentang penundaan pemilu 2014 dengan menggunakan tool rapidminer dengan tambahan services *Twiter* Data API. Setelah dilakukan seleksi, didapatkan sebanyak 1.826 komentar yang sesuai dengan penelitian. Dataset dilabeli secara manual berupa kalimat positif dan negatif menggunakan dengan bantuan orang lain sebanyak 700 data sebagai *data training*, lalu dilakukan analisis data secara otomatis menggunakan *data testing*.

#### 3) Data Preparation

Pada tahapan data preparation ini maka data yang sudah terkumpul akan dilakukan *data preprocessing* oleh peneliti. Berikut ini adalah langkah-langkah yang dilakukan selama data preprocessing :

- Remove duplicates*, disini peneliti melakukan proses remove duplicates untuk menghapus data yang muncul secara berulang.
- Case folding*, disini peneliti melakukan proses *case folding* dengan menggunakan operator *transform cases* pada rapidminer untuk mengubah semua data komentar menjadi huruf kecil.
- Tokenizing*, disini peneliti melakukan proses tokenizing untuk memecah komentar dari kalimat menjadi perkata, menghilangkan karakter yang bersifat pembatas, dan menghapus angka atau tanda baca.
- Stopwords*, disini peneliti melakukan *stopwords* untuk menghilangkan kata yang tidak penting atau tidak deksriptif dengan menggunakan *stopword list* dari Tala.



Gambar 2. Preprocessing Data

#### 4) Modelling

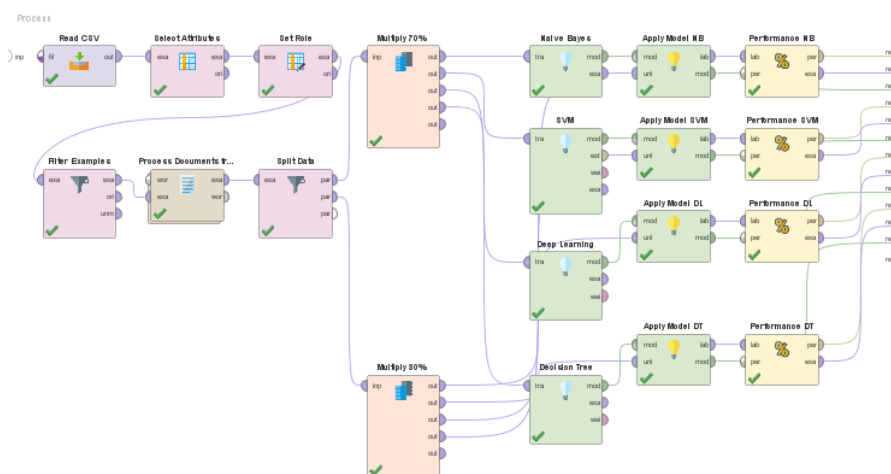
Pada tahapan modelling ini, maka peneliti akan melakukan pengukuran performa klasifikasi dalam dua tahap :

a) *Modelling* menggunakan *split data*

Pada tahapan ini maka data akan di *split data* menggunakan model 80:20, data dilatih menggunakan algoritma NB, SVM, DL, dan DT. Kemudian pada tahap testing, data diproses melalui Apply Model dan Performance Test

Tabel 1. Evaluasi Accuracy Berdasarkan Model *Train/Test Split Data*

Algoritma	Akurasi	Precision	Recall
Naive Bayes	80.55%	72.05%	81.69%
Support Vektor Machine	76.44%	81.11%	51.41%
Deep Learning	81,37%	81.90%	66.90%
Decision Tree	63.29%	75.00%	8.45%



Gambar 3. *Train/Test Split Data*

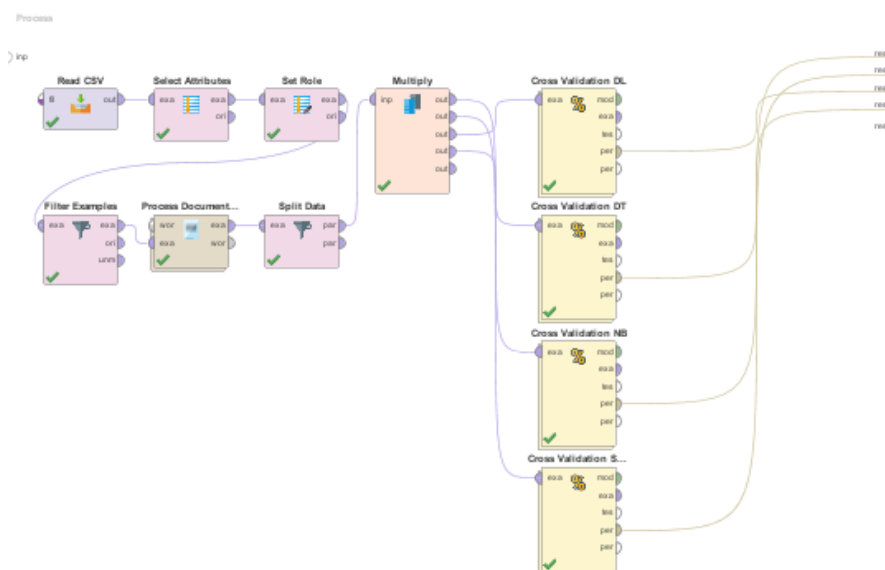
Berdasarkan Tabel 1 maka disimpulkan bahwa akurasi tertinggi sebesar 81,37% pada algoritma *deep learning*, yang dihasilkan dengan pemodelan terhadap *dataset* yang sudah dilakukan *preprocessing* menggunakan perbandingan 70:30 untuk *data training* dan *data testing*.

b) *Modelling* menggunakan *cross validation*

Pada tahapan ini maka *data training* dan *data testing* di *split* secara acak menjadi 10 bagian ( $k = 10$ )

Tabel 2. Evaluasi Accuracy Berdasarkan Model *K-Fold Cross Validation*

Algoritma	Akurasi	Precision	Recall
Naive Bayes	79.81%	70.26%	83.85%
Support Vektor Machine	77.34%	85.08%	51.21%
Deep Learning	77.89%	72.47%	70.09%
Decision Tree	64.07%	83.69%	9.71%



Gambar 4. K-Fold Cross Validation

Berdasarkan Tabel 1 dan Tabel 2 yang berisi hasil pengukuran performa model klasifikasi *naive bayes*, *support vektor machine*, *deep learning* dan *decision tree* dengan 2 cara yaitu pembagian *dataset* dengan *split data* dan *cross validation*, maka dapat disimpulkan bahwa pemodelan dengan *split data* 80:20 memiliki akurasi sebesar 81,37% algoritma *deep learning* dan pemodelan dengan *cross validation* memiliki akurasi sebesar 77,89% sehingga akurasi tertinggi yang dihasilkan pada penelitian ini adalah dengan pemodelan menggunakan *split data* 80:20.

#### 5) Evaluation

Penelitian ini akan melakukan pengujian dengan menggunakan *confussion matrix* yang diperoleh dari tahapan *modelling* dengan menggunakan algoritma *naive bayes*, *support vektor machine*, *deep learning* dan *decision tree* akurasi dengan *split data* 80:20 lebih tinggi dibandingkan *cross validation*. Sehingga *confussion matrix* yang disajikan pada Tabel 3 adalah hasil dari evaluasi pengukuran klasifikasi *deep learning* dengan *dataset* yang berjumlah 1826, terdapat 1126 data training, dan 700 data testing.

Tabel 3. Confusion Matrix Algoritma Deep Learning

	True Positive	True Negatif	Class Precision
Pre. Positive	182	34	84.26
Pre. Negative	41	108	72.48
Class Recall	81.61%	76.06 %	

Perhitungan dari Tabel 3 adalah sebagai berikut :

- Precision =  $\frac{TP}{TP+FP} = \frac{182}{182+34} = 0.8426 = 84.26\%$
- Recall =  $\frac{TP}{TP+FN} = \frac{182}{182+41} = 0.8161 = 81.61\%$
- Accuracy =  $\frac{TP}{TP+TN+FP+FN} = \frac{182+108}{182+108+34+41} = 0.8579 = 85.80\%$

## 4. Kesimpulan

Berdasarkan hasil komentar penundaan pemilu 2024 yang didapat setelah melakukan penelitian ini, maka dapat diambil kesimpulan bahwa berdasarkan proses data hasil crawling dari media sosial



twitter dan klasifikasi model dengan algoritma naive bayes, support vektor machine, deep learning dan decision tree, menggunakan model split data dengan perbandingan 80:20 terhadap dataset komentar pengguna pada rapidminer, maka penggunaan algoritma deep learning menghasilkan nilai akurasi lebih baik dibandingkan dengan algoritma NB, SVM dan DT. Sehingga algoritma deep learning yang menggunakan model train/ test split data menjadi solusi untuk melakukan klasifikasi analisis sentimen masyarakat Indonesia terhadap penundaan pemilu 2024.

Dari kesimpulan diatas dapat diberikan saran-saran sebagai berikut :

- 1) Penulis sangat menyarankan agar dilakukan penelitian atau pengembangan lebih lanjut terhadap penelitian ini dengan metode lainnya dan juga menggunakan dataset yang lebih baru dan lebih banyak dari penelitian sebelumnya.
- 2) Hasil dari proses analisis sentiment ini dapat digunakan sebagai pertimbangan masyarakat dalam mengambil keputusan lebih lanjut dalam penelitian selanjutnya agar proses *crawling* tidak hanya dilakukan pada media sosial *twitter* dan optimalisasi algoritma dengan menggunakan algoritma klasifikasi berbasis *Particle Swarm Optimization* agar dapat dibandingkan dan mendapat akurasi yang paling paling baik.

## 5. Daftar Pustaka

- [1] M. Ikbali, 2022. *Apakah pemilihan umum 2024 tetap dilaksanakan atau ditunda*. URL: <https://stialan.ac.id/v3/apakah-pemilihan-umum-2024-tetap-dilaksanakan-atau-ditunda/>
- [2] D. Nurita, 2022. *Busyro Muqoddas Sebut Pengusul Tunda Pemilu 2024 Tak Punya Malu*. URL: <https://nasional.tempo.co/read/1571423/busyro-muqoddas-sebut-pengusul-tunda-pemilu-2024-tak-punya-malu>
- [3] Amirulloh, 2022. *Denny Indrayana: Usul Penundaan Pemilu 2024 Pelanggaran Konstitusi Berjamaah*. URL: <https://nasional.tempo.co/read/1564734/denny-indrayana-usul-penundaan-pemilu-2024-pelanggaran-konstitusi-berjamaah>
- [4] D. Nurita, 2022. *Pakar Hukum UGM Khawatirkan Dua Skenario yang Bisa Loloskan Penundaan Pemilu*. URL: <https://nasional.tempo.co/read/1571591/pakar-hukum-ugm-khawatirkan-dua-skenario-yang-bisa-loloskan-penundaan-pemilu>
- [5] S. Juariya, 2022. *Dampak yang Akan Terjadi Jika Pelaksanaan Pilkada Serentak Tahun 2024 Konten ini telah tayang di Kompasiana.com dengan judul "Dampak yang Akan Terjadi Jika Pelaksanaan Pilkada Serentak Tahun 2024", Klik untuk baca: https://www.kompasiana.com/sitijuariya72*. URL: <https://www.kompasiana.com/sitijuariya7226/62a3340efca4e4528d13bd43/dampak-yang-akan-terjadi-jika-pelaksanaan-pilkada-serentak-tahun-2024>
- [6] C. M. Annur, 2022. *Pengguna Twitter Indonesia Masuk Daftar Terbanyak di Dunia, Urutan Berapa?*. URL: <https://databoks.katadata.co.id/datapublish/2022/03/23/pengguna-twitter-indonesia-masuk-daftar-terbanyak-di-dunia-urutan-berapa>
- [7] Muhammad Fadhl, 2022. *Analisa Sentimen Mengenai Pendidikan Tatap Muka pada Twitter Menggunakan Metode Naïve Bayes dan Support Vector Machine*, Dr. Diss. Univ. Mercu Buana Jakarta.
- [8] Saputro, I.W. and Sari, B.W., 2020. Uji Performa Algoritma Naïve Bayes untuk Prediksi Masa Studi Mahasiswa. *Creative Information Technology Journal*, 6(1), pp.1-11. DOI: <https://doi.org/10.24076/citec.2019v6i1.178>.

- [9] Nida, E.A., 2020. Analisis Kinerja Algoritma Support Vector Machine (SVM) Guna Pengambilan Keputusan Beli/Jual Pada Saham PT Elnusa Tbk.(ELSA). *Jurnal Transformatika*, 17(2), pp.160-170. DOI: <http://dx.doi.org/10.26623/transformatika.v17i2.1649>.
- [10] Sari, R., 2020. Analisis sentimen pada review objek wisata dunia fantasi menggunakan algoritma K-Nearest Neighbor (k-nn). *EVOLUSI: Jurnal Sains Dan Manajemen*, 8(1). DOI: <https://doi.org/10.31294/evolusi.v8i1.7371>.
- [11] Harun, A. and Ananda, D.P., 2021. Analisa Sentimen Opini Publik Tentang Vaksinasi Covid-19 di Indonesia Menggunakan Naïve bayes dan Decision Tree: Analysis of Public Opinion Sentiment About Covid-19 Vaccination in Indonesia Using Naïve Bayes and Decision Tree. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 1(1), pp.58-64. DOI: <https://doi.org/10.57152/malcom.v1i1.63>.
- [12] Sihombing, R.E., Rachmatin, D. and Dahlan, J.A., 2019. Program Aplikasi Bahasa R Untuk Pengelompokan Objek Menggunakan Metode K-Medoids Clustering. *Jurnal EurekaMatika*, 7(1), pp.58-79.
- [13] Ardika, R.B.P., Irawan, B. and Setianingsih, C., 2020. Analisis Sentimen Data Pada Bpjs Kesehatan Menggunakan Backpropagation Neural Network. *eProceedings of Engineering*, 7(2).
- [14] Yang, L., 2020. Wang and Sherratt, 2020 Yang Li, Li Y., Wang J., Sherratt RS. *Sentiment analysis for E-commerce product reviews in Chinese based on sentiment lexicon and deep learning*, *IEEE Access*, 8, pp.23522-23530. DOI: <https://doi.org/10.1109/ACCESS.2020.2969854>.
- [15] Pratiwi, A.R.D. and Setiawan, E.B., 2020. Implementation of rumor detection on twitter using the svm classification method. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 4(5), pp.782-789. DOI: <https://doi.org/10.29207/resti.v4i5.2031>.
- [16] Aletti, G., Crimaldi, I. and Saracco, F., 2021. A model for the Twitter sentiment curve. *Plos one*, 16(4), p.e0249634. DOI: <https://doi.org/10.1371/journal.pone.0249634>.
- [17] Chandani, V., Wahono, R.S. and Purwanto, P., 2015. Komparasi algoritma klasifikasi Machine Learning dan feature selection pada analisis sentimen review film. *Journal of Intelligent Systems*, 1(1), pp.56-60.
- [18] Putra, A.D.A. and Juanita, S., 2021. Analisis Sentimen pada Ulasan pengguna Aplikasi Bibit Dan Bareksa dengan Algoritma KNN. *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, 8(2), pp.636-646. DOI: <https://doi.org/10.35957/jatisi.v8i2.962>.
- [19] Harsono, M.L.K., Alkhalifi, Y. and Gata, W., 2020. Analisis Sentimen Stakeholder Atas Layanan HAIDJPB Pada Media Sosial Twitter Dengan Menggunakan Metode Support Vector Machine Dan Naïve Bayes. *Infoman's: Jurnal Ilmu-ilmu Manajemen dan Informatika*, 14(1).
- [20] Melita, R., Amrizal, V., Suseno, H.B., Dirjam, T., Informatika, T. and Sains, F., 2018. Penerapan Metode Term Frequency Inverse Document Frequency (Tf-Idf) Dan Cosine Similarity Pada Sistem Temu Kembali Informasi Untuk Mengetahui Syarah Hadits Berbasis Web (Studi Kasus: Syarah Umdatil Ahkam). *J. Tek. Inform*, 11(2), pp.149-164.