DOI: https://doi.org/10.37817/ikraith-informatika.v9i3
P-ISSN : 2580-4316
E-ISSN : 2654-8054

PERBANDINGAN METODE MACHINE LEARNING UNTUK ANALISIS SENTIMEN OPINI PUBLIK TENTANG TAPERA (TABUNGAN PERUMAHAN RAKYAT)

[1]Gery Octavansyah, [2]Bahrul Ulum
[1,2]Teknik Informatika, Fakultas Ilmu Komputer, Universitas Esa Unggul, DKI Jakarta

E-mail: [1]geryoctavansyah17@student.esaunggul.ac.id
[2]m.bahrul_ulum@esaunggul.ac.id

ABSTRAK

Kebijakan Tabungan Perumahan Rakyat (TAPERA) telah memicu perdebatan luas di masyarakat Indonesia, terutama di platform media sosial X (sebelumnya Twitter). Penelitian ini bertujuan untuk menganalisis sentimen opini publik terhadap implementasi TAPERA dengan membandingkan berbagai metode *Machine Learning* dan *Deep Learning*. Data dikumpulkan menggunakan teknik *web scraping* melalui platform Apify dalam rentang waktu Mei hingga Juli 2024, menghasilkan total 5.580 tweet. Sebelum tahap klasifikasi, data melalui proses *preprocessing* menggunakan teknik *Natural Language Processing* (NLP) untuk menangani variasi bahasa informal, *slang*, dan penghapusan elemen non-tekstual. Hasil analisis menunjukkan bahwa mayoritas opini publik terhadap TAPERA bersifat negatif. Berdasarkan pengujian performa, model *Decision Tree* dan *Support Vector Machine* (SVM) memberikan tingkat akurasi tertinggi masing-masing sebesar 0,99 pada *Confusion Matrix*. Sementara itu, model *Deep Learning* LSTM memberikan hasil akurasi sebesar 0,5624. Penelitian ini memberikan kontribusi praktis bagi pemangku kepentingan dalam memahami dinamika opini publik sebagai dasar pengambilan keputusan strategis terkait kebijakan perumahan nasional.

Kata kunci :*Analisis Sentimen, TAPERA, Machine Learning, Deep Learning, Media Sosial X.*

ABSTRACT

The People's Housing Savings Policy (TAPERA) has sparked widespread debate in Indonesian society, especially on the social media platform X (formerly Twitter). This study aims to analyze the sentiment of public opinion on the implementation of TAPERA by comparing various *Machine Learning* and *Deep Learning* methods. The data was collected using *web scraping techniques* through the Apify platform in the period from May to July 2024, resulting in a total of 5,580 tweets. Before the classification stage, the data goes through *a preprocessing* process using *Natural Language Processing* (NLP) techniques to handle informal language variations, *slang*, and the removal of non-textual elements. The results of the analysis show that the majority of public opinion towards TAPERA is negative. Based on performance testing, *the Decision Tree* and *Support Vector Machine* (SVM) models provided the highest accuracy level of 0.99 each on *the Confusion Matrix*. Meanwhile, the LSTM Deep Learning model gave an accuracy of 0.5624. This research provides practical contributions for stakeholders in understanding the dynamics of public opinion as the basis for strategic decision-making related to national housing policy.

Keyword : ***Analisis Sentimen, TAPERA, Machine Learning, Deep Learning, Media Sosial X***

P-ISSN : 2580-4316
E-ISSN : 2654-8054 DOI: https://doi.org/10.37817/ikraith-informatika.v9i3

1. PENDAHULUAN

Kebijakan Tabungan Perumahan Rakyat (TAPERA) yang diluncurkan oleh Pemerintah Indonesia bertujuan untuk menghimpun dan menyediakan dana murah jangka panjang yang berkelanjutan untuk pembiayaan perumahan bagi masyarakat. Meskipun memiliki tujuan mulia untuk mengatasi *backlog* perumahan, implementasi kebijakan ini memicu pro dan kontra yang masif di tengah masyarakat (Palanisamy et al., 2013). Hal ini terlihat dari tingginya volume diskusi di platform media sosial, khususnya X (sebelumnya Twitter), yang menjadi saluran utama masyarakat dalam menyampaikan aspirasi, keluhan, maupun dukungan secara real-time.

Opini publik di media sosial merupakan data tekstual yang sangat berharga namun memiliki volume yang sangat besar dan tidak terstruktur. Untuk memahami kecenderungan suara masyarakat secara efektif, diperlukan pendekatan komputasional melalui Analisis Sentimen. Namun, tantangan utama dalam analisis sentimen bahasa Indonesia di media sosial adalah penggunaan bahasa yang tidak baku, singkatan, (Recupero et al., 2014).serta istilah-istilah gaul (*slang*) yang dapat mempengaruhi performa klasifikasi.

Beberapa penelitian terdahulu telah menggunakan berbagai algoritma *Machine Learning* untuk klasifikasi sentimen, seperti *Support Vector Machine* (SVM), *Decision Tree*, dan *Naive Bayes*. (Hutto & Gilbert, 2014) Di sisi lain, pendekatan *Deep Learning* seperti *Long Short-Term Memory* (LSTM) juga mulai populer karena kemampuannya dalam menangani konteks urutan kata dalam kalimat. Meskipun demikian, efektivitas masing-masing algoritma seringkali bervariasi tergantung pada karakteristik data dan kasus yang diangkat.

Penelitian ini bertujuan untuk melakukan analisis sentimen terhadap opini publik mengenai kebijakan TAPERA dengan membandingkan performa beberapa metode *Machine Learning* (SVM dan *Decision Tree*) dan *Deep Learning* (LSTM). Dengan menggunakan data yang dikumpulkan dari platform X selama periode puncak perdebatan (Gurkhe et al., 2014), penelitian ini diharapkan dapat memberikan gambaran objektif mengenai respons masyarakat. Selain itu, perbandingan model yang dilakukan bertujuan untuk menemukan algoritma yang paling akurat dan efisien dalam mengklasifikasikan sentimen terkait kebijakan publik di Indonesia. Hasil dari penelitian ini diharapkan dapat menjadi masukan bagi pemerintah sebagai bahan evaluasi dalam komunikasi kebijakan serta pengembangan strategi nasional di masa mendatang.

2. LANDASAN TEORI

2.1 TAPERA (Tabungan Perumahan Rakyat)

TAPERA adalah program pemerintah Indonesia berdasarkan Undang-Undang Nomor 4 Tahun 2016 yang dirancang untuk membantu masyarakat, khususnya berpenghasilan rendah dan menengah, dalam memperoleh rumah layak huni. Program ini berfungsi sebagai mekanisme pembiayaan perumahan melalui pengumpulan dan pengelolaan dana wajib dari peserta, baik pekerja formal maupun informal. Tujuan utamanya adalah meningkatkan aksesibilitas hunian terjangkau dan mengurangi *backlog* perumahan nasional.

2.2 Media Sosial

Media sosial adalah platform digital yang memungkinkan pengguna untuk membuat, berbagi, dan bertukar informasi, ide, dan konten dalam bentuk teks, gambar, *video*, dan *audio*. Media sosial mencakup berbagai jenis platform, termasuk jaringan sosial (seperti *Facebook* dan *LinkedIn*), situs berbagi media (seperti *YouTube* dan *Instagram*), serta platform mikroblogging (seperti X). (Boyd) mendefinisikan media sosial sebagai layanan berbasis web yang memungkinkan individiu untuk membangun profil publik atau semi-

DOI: https://doi.org/10.37817/ikraith-informatika.v9i3
P-ISSN : 2580-4316
E-ISSN : 2654-8054

publik dalam suatu sistem terbatas, mengartikan daftar pengguna lain yang berbagi koneksi, dan melihat serta melintasi daftar koneksi mereka sendiri dan yang dibuat oleh orang lain dalam sistem tersebut.

2.3 Machine Learning

Machine Learning (ML) adalah cabang dari kecerdasan buatan (*Artificial Intelligence*) yang berfokus pada pengembangan algoritma dan teknik yang memungkinkan komputer untuk "belajar" dari dan membuat prediksi atau keputusan berdasarkan data. ML adalah proses di mana sistem komputer secara otomatis memperbaiki kinerjanya melalui pengalaman tanpa perlu diprogram secara eksplisit. Konsep dasar ML pertama kali diperkenalkan oleh Arthur Samuel pada tahun 1959, yang mendefinisikannya sebagai "*the field of study that gives computers the ability to learn without being explicitly programmed*" (Samuel, 1959).

2.4 Confusion Matrix

Kemampuan Confusion Matrix adalah alat evaluasi yang digunakan dalam masalah klasifikasi untuk menilai kinerja model klasifikasi. Matriks ini memberikan gambaran rinci tentang bagaimana model membuat prediksi, menunjukkan jumlah prediksi yang benar dan salah untuk setiap kelas dalam *dataset*. Confusion Matrix digunakan untuk menghitung berbagai metrik evaluasi seperti akurasi, presisi, *recall*, dan *F1-score* (Fawcett, 2006). Confusion Matrix adalah tabel yang menggambarkan performa model klasifikasi dalam hal prediksi positif dan negatif untuk setiap kelas.

2.5 Cross-Validation

Cross-validation adalah teknik evaluasi model yang digunakan untuk mengukur kinerja prediksi dari model statistik atau machine learning pada dataset independen yang tidak terlihat selama proses pelatihan. Metode ini bertujuan untuk menghindari overfitting, memberikan estimasi performa model yang lebih stabil, dan membantu dalam pemilihan model yang optimal (Kohavi, 1995).

2.6 Natural Language Processing (NLP)

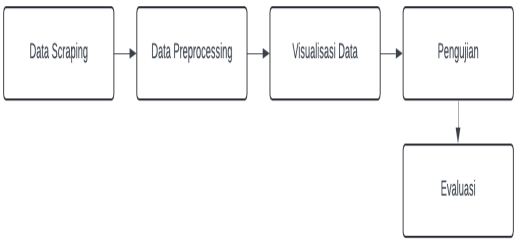
Natural Language Processing (NLP) adalah bidang studi dalam kecerdasan buatan (AI) dan linguistik komputer yang berfokus pada interaksi antara komputer dan bahasa alami manusia. NLP bertujuan untuk memungkinkan komputer memahami, menafsirkan, dan menghasilkan teks atau ucapan dalam bahasa yang digunakan manusia. Bidang ini mencakup berbagai teknik untuk analisis teks, pengenalan suara, terjemahan bahasa, dan banyak aplikasi lainnya (Jurafsky & Martin, 2024).

3. METODOLOGI

Penelitian ini menganalisis sentimen publik terkait kebijakan TAPERA menggunakan data dari media sosial X yang dikumpulkan melalui platform Apify antara Mei hingga Juli 2024. Data diambil secara berkala setiap 10 hingga 15 menit dengan mengunduh 100 tweet per sesi hingga mencapai target 5.000 data menggunakan kueri spesifik seperti #BPTapera dan "TAPERA". Tahap awal penelitian dimulai dengan *data scraping* untuk memperoleh data mentah berdasarkan kata kunci dan rentang waktu yang ditentukan. Selanjutnya, data melalui proses *preprocessing* untuk ditransformasikan ke dalam format yang siap diolah melalui teknik tokenisasi, *stemming*, dan *lemmatization*. Identifikasi sentimen dilakukan dengan mengklasifikasikan ekspresi publik ke dalam kategori positif, negatif, dan netral menggunakan model berbasis leksikon InSet dan algoritma pembelajaran mesin. Hasil analisis kemudian divisualisasikan menggunakan alat bantu seperti WordCloud dan Seaborn untuk menggambarkan pola distribusi sentimen masyarakat secara jelas. Pada tahap akhir, dilakukan pengujian model dengan membandingkan metode Naïve Bayes, Logistic Regression, dan Long-Short Term Memory (LSTM) yang kemudian dievaluasi akurasinya

P-ISSN : 2580-4316
E-ISSN : 2654-8054 DOI: https://doi.org/10.37817/ikraith-informatika.v9i3

melalui perhitungan *confusion matrix* serta *cross-validation*.



Gambar 1. Tahapan Penelitian

3.1 Teknik Pengumpulan Data

3.1.1 Identifikasi Tagar dan Kata Kunci

Tahap Langkah pertama dalam pengumpulan data adalah mengidentifikasi tagar (*#hashtag*) dan kata kunci yang relevan dengan TAPERA. Beberapa tagar dan kata kunci yang mungkin digunakan termasuk:

a. Tagar:#Tapera,#TabunganPerumahan Rakyat, #BPTapera
b. Kata Kunci: "TAPERA Tabungan Perumahan Rakyat", "BPTAPERA"

3.1.2 Penggunaan Apify

Untuk mengumpulkan *tweet* yang relevan, Apify akan digunakan. Apify memungkinkan peneliti untuk mengunduh *tweet* berdasarkan kriteria tertentu, seperti tagar, kata kunci, dan rentang waktu. Langkah-langkah spesifik untuk penggunaan Apify adalah sebagai berikut:

1. **Pemasukkan Kata Kunci dan Tagar**
   Memasukkan kata kunci dan tagar yang digunakan
2. **Pemfilteran Data**
   Data di-*filter* berdasarkan bahasa dan tanggal
3. **Memulai Pengambilan Data**
   Data diambil sebanyak 100 data *per run*.
4. **Export Data**
   Data di-*export* dengan mengambil kolom ‘id’ dan ‘text’.

3.1.3 Pemilihan Periode Waktu

Periode waktu pengumpulan data sangat penting untuk menangkap dinamika sentimen. Periode waktu yang akan dipilih yaitu dari Mei 2024 hingga Juli 2024.

3.1.4 Alat Bantu Penelitian

Tabel 1. Alat Bantu Penelitian

No	Nama	Jenis	Fungsi
1	Google Colab	Software	Menulis kode
2	Microsoft Word	Software	Menulis naskah proposal tugas akhir dan dokumentasi
3	Microsoft Excel	Software	Menyimpan dan mengolah data
4	LucidApp	Software	Membuat diagram-diagram
5	Apify	Software	Mengambil data dari sosial media X
6	Pandas	Library	Manipulasi data dan analisis data
7	NumPy	Library	Memproses operasi-operasi numerikal
8	NLTK	Library	Memproses bahasa alami
9	Sastrawi	Library	Memproses bahasa alami
10	WordCloud	Library	Visualisasi data tekstual
11	Seaborn	Library	Visualisasi data numerikal
12	MatPlotLib	Library	Visualisasi data numerikal
13	Datetime	Library	Memproses tanggal dan waktu
14	Regex	Library	Manipulasi string
15	IndoNLP	Library	Data Preprocessing

DOI: https://doi.org/10.37817/ikraith-informatika.v9i3
P-ISSN : 2580-4316
E-ISSN : 2654-8054

4. HASIL DAN PEMBAHASAN

4.1 Data Hasil Penelitian

Pada Bab ini Proses pengolahan data dimulai dengan pengambilan sebanyak 5.580 tweet dari media sosial X menggunakan aktor *Tweet Scraper v2* pada platform Apify dalam rentang waktu Mei hingga Juli 2024. Data mentah tersebut kemudian diproses melalui jalur *preprocessing* yang mencakup penghapusan elemen HTML, URL, angka, serta normalisasi bahasa *slang* dan konversi emoji menjadi teks menggunakan pustaka IndoNLP. Hasil visualisasi distribusi sentimen menggunakan pustaka Seaborn menunjukkan bahwa opini publik mengenai kebijakan TAPERA didominasi secara signifikan oleh sentimen negatif dibandingkan sentimen positif dan netral. Analisis tematik melalui *WordCloud* mengidentifikasi bahwa kata-kata yang paling sering muncul dalam percakapan publik meliputi "tapera", "potong", "gaji", dan "rakyat", yang mencerminkan kekhawatiran masyarakat terhadap pemotongan pendapatan.

Pada tahap klasifikasi, penelitian ini membandingkan performa berbagai algoritma *machine learning* dan *deep learning*. Hasil pengujian menunjukkan bahwa model *Decision Tree*, *Support Vector Machine* (SVM), dan *Logistic Regression* memberikan performa tertinggi dengan nilai akurasi pada *Confusion Matrix* mencapai 0,99. Model *Random Forest* juga menunjukkan hasil yang sangat baik dengan akurasi 0,99 dan skor *Cross-Validation* sebesar 0,97. Sementara itu, model *Naive Bayes* memperoleh akurasi sebesar 0,93 dengan skor *Cross-Validation* 0,92. Di sisi lain, pengujian menggunakan model *deep learning* FastText menghasilkan akurasi sebesar 0,99. Namun, model LSTM memberikan hasil yang lebih rendah dibandingkan model lainnya, dengan akurasi rata-rata tertinggi sebesar 0,5624 yang dicapai menggunakan parameter *Dropout* 0,4 pada 10 *epoch*. Hal ini mengindikasikan bahwa untuk karakteristik dataset teks bahasa Indonesia terkait kebijakan publik ini, model pembelajaran mesin konvensional memiliki efektivitas yang lebih tinggi dalam melakukan klasifikasi dibandingkan arsitektur saraf yang lebih kompleks.

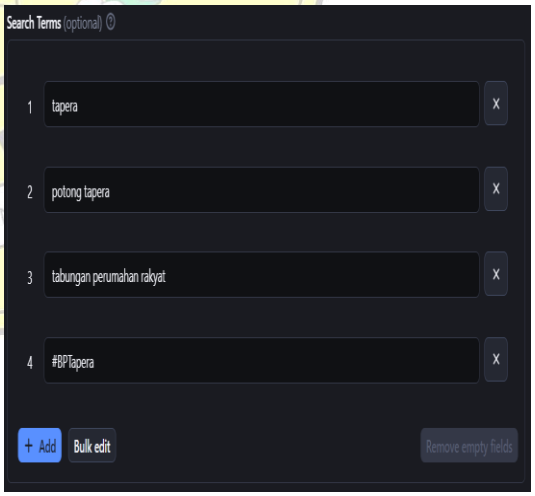
4.1.1 Teknik *Data Scraping*

Data Scraping dilakukan menggunakan *tool* Apify dengan actor "*Tweet Scaper v2*". Langkah yang dilakukan untuk proses *Data Scraping* adalah sebagai berikut:

1. Input *Keywords*

   Agar hasil data scraping sesuai dengan topik, maka diperlukan *keywords* yang berguna untuk mem-*filter* data yang diambil agar sesuai. *Keywords* yang digunakan adalah:

- Tapera
- Potong Tapera
- Tabungan Perumahan Rakyat
- #BPTapera

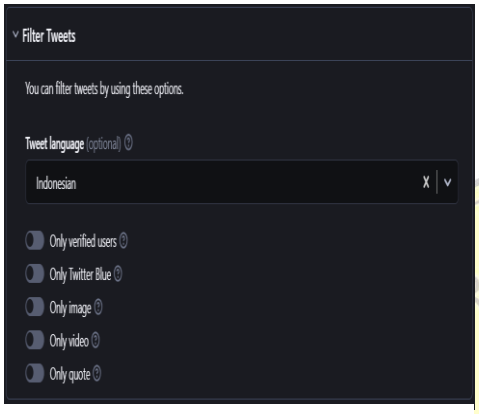


Gambar 2. Input Keyword

P-ISSN : 2580-4316
E-ISSN : 2654-8054 DOI: https://doi.org/10.37817/ikraith-informatika.v9i3

4.1.2 Filter by Language

Dikarenakan beberapa bahasa memiliki kata “tapera” sebagai bahasa slang yang digunakan, maka diperlukan untuk mem-filter data sesuai dengan bahasa yang diinginkan, yaitu Bahasa Indonesia.



Gambar 3. Filter by Language

4.1.3 Filter by Date

Pengambilan data dilakukan sesuai dengan jangka waktu yang ditentukan agar data-data yang diambil tetap relevan.

Gambar 4. Filter by Date

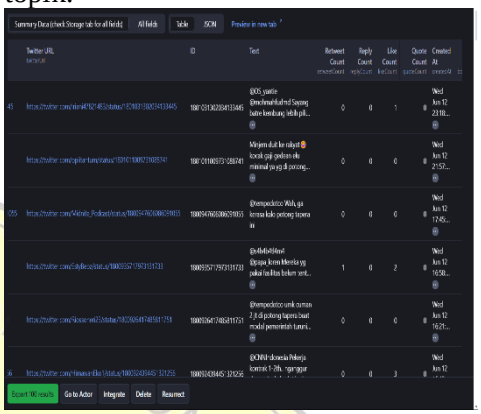
4.1.4 Start Scraper

Setelah input-input seperti filter dan keywords sudah ditambahkan, maka dapat dilakukan pengambilan data.

Gambar 5. Start Scraper

4.1.5 Check data

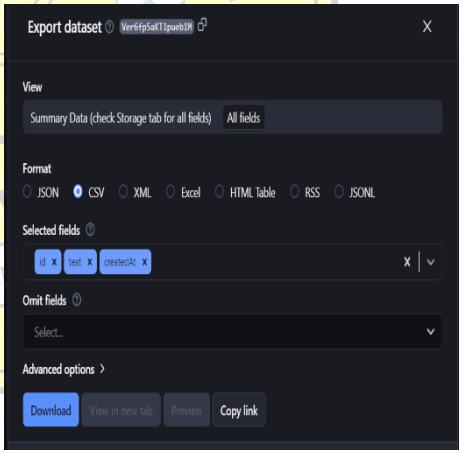
Data yang diambil sebanyak 100 data per sesi pengambilan data. Pengecekan dapat dilakukan agar data-data yang terambil relevan dan sesuai dengan topik.



Gambar 6. Check Data

4.1.6 Export Data

Data dapat di-export agar dapat digunakan dalam program. Kolom yang diambil adalah kolom id, text, dan createdAt



Gambar 7. Export Data

DOI: https://doi.org/10.37817/ikraith-informatika.v9i3
P-ISSN : 2580-4316
E-ISSN : 2654-8054

4.2 Pembahasan Hasil Penelitian

4.2.1 Import Libraries

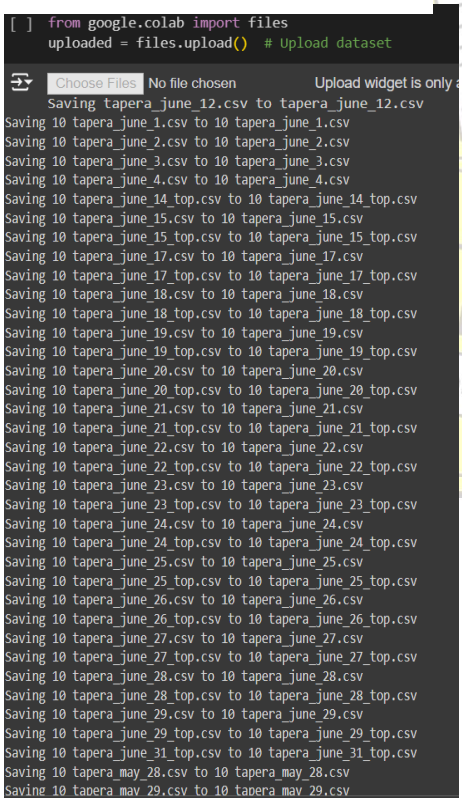
Memasukkan dan meng-install library-library yang digunakan pada program.

Gambar 8. Import Liraries

4.2.2 Import Data

Memasukkan data yang sudah diambil ke dalam program. Data yang dimasukkan adalah data text dari sosial media X dan dataset lexicon negatif dan positif untuk sentimen analisis.

Gambar 10. Import Data



Dataset lalu akan di-load dan ditambahkan kolom "sentiment" untuk pelabelan data. Kolom sentimen ini nantinya akan diisi oleh label sentimen (positif, negatif, atau netral) pada tahap sentimen analisis.

Gambar 9. Load DataSet

4.2.3 Data Preprocessing

Data preprocessing dilakukan menggunakan library IndoNLP dan RegEx.

```
# Create the preprocessing pipeline
preprocessing_pipeline = pipeline([
    remove_html,
    remove_url,
    replace_word_elongation,
    replace_slang,
    emoji_to_words,
    remove_stopwords
])

# Preprocess the text using indoNLP
def preprocess_text(text):
    text = preprocessing_pipeline(text)  # Apply the preprocessing pipeline
    text = re.sub(r'@\w+', '', text)     # Remove mentions
    text = re.sub(r'#\w+', '', text)     # Remove hashtags
    text = re.sub(r'\d+', '', text)      # Remove numbers
    text = text.lower()                  # Convert to lowercase
    tokens = word_tokenize(text)         # Tokenize the text
    return ' '.join(tokens)
```

P-ISSN : 2580-4316
E-ISSN : 2654-8054 DOI: https://doi.org/10.37817/ikraith-informatika.v9i3

Gambar 11. Data Preprocessing

4.2.4 Sentimen Analisis

Data-data akan dihitung polarity-nya, lalu dilakukan analisis sentimen berdasarkan jumlah polarity dari data.

```
# Load sentiment lexicons from the provided .tsv files
positive_words = pd.read_csv('positive.tsv', sep='\t', header=None)[0].tolist()
negative_words = pd.read_csv('negative.tsv', sep='\t', header=None)[0].tolist()

# Classify sentiment using the lexicons
def classify_sentiment(text):
    tokens = word_tokenize(text)
    positive_count = sum(1 for word in tokens if word in positive_words)
    negative_count = sum(1 for word in tokens if word in negative_words)
    if positive_count > negative_count:
        return "positive"
    elif negative_count > positive_count:
        return "negative"
    else:
        return "neutral"
```

Gambar 12. Sentimen Analisis

Pada Gambar 12, terlihat proses sentimen analisis menggunakan leksikon. Tahap pertama yaitu memasukkan leksikon positif dan leksikon negatif (positive.tsv dan negative.tsv), lalu me-load leksikon tersebut menggunakan library pandas.

Tabel 2. Contoh kata pada leksikon positif

No	Kata	Polarity
1	menyukai	4
2	rajin	4
3	gembira	5
4	luar biasa	4
5	terimakasih	5
6	antusias	2
7	menyala	2
8	benar	3
9	kawal	1
10	teladan	4

Tabel 3. Contoh kata pada leksikon negatif

No	Kata	Polarity
1	sakit	-5
2	gamau	-4
3	nggak	-3
4	males	-4
5	gaada	-3
6	beber	-2
7	gila	-4
8	serem	-5
9	bosen	-4
10	pusing	-4

4.2.5 Long Short Term Memory (LSTM)

LSTM akan diuji menggunakan 10 Epoch dengan beberapa kombinasi Dropout (dari 0.1 hingga 0.9), Hasil akan ditampilkan dalam tabel dan screenshot

Tabel 4. Long ShortTerm Memory (LSTM)

Epoch	0.1	0.2	0.3	0.4
1	0.5740	0.5360	0.5541	0.5796
2	0.5579	0.5541	0.5576	0.5647
3	0.5676	0.5581	0.5679	0.5524
4	0.5712	0.5681	0.5658	0.5680
5	0.5524	0.5554	0.5649	0.5610
6	0.5578	0.5523	0.5551	0.5596
7	0.5493	0.5614	0.5718	0.5629
8	0.5571	0.5758	0.5610	0.5480
9	0.5602	0.5533	0.5658	0.5594
10	0.5630	0.5523	0.5586	0.5693

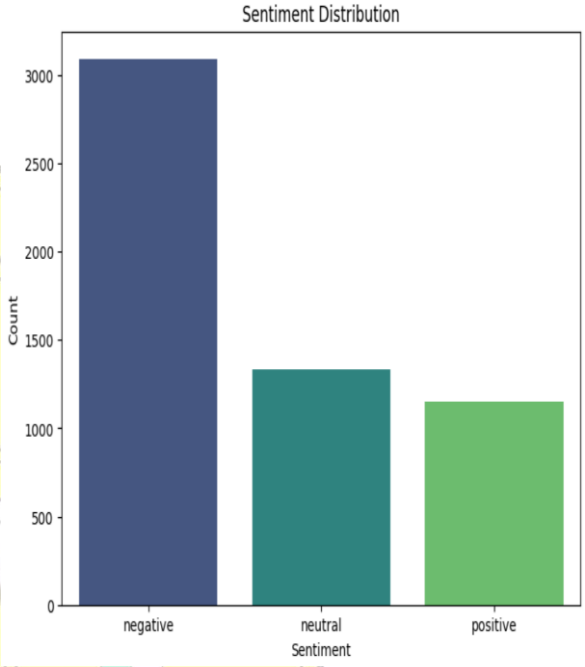
Epoch	0.5	0.6	0.7	0.8	0.9

DOI: https://doi.org/10.37817/ikraith-informatika.v9i3
P-ISSN : 2580-4316
E-ISSN : 2654-8054

1	0.54 87	0.56 08	0.55 92	0.55 96	0. 54 30
2	0.56 03	0.55 83	0.56 19	0.55 83	0. 56 29
3	0.56 06	0.56 06	0.56 68	0.54 83	0. 56 51
4	0.55 98	0.56 29	0.57 07	0.55 66	0. 55 84
5	0.55 32	0.56 29	0.55 33	0.56 16	0. 56 82
6	0.55 33	0.56 11	0.54 97	0.56 03	0. 56 46
7	0.55 45	0.55 35	0.56 09	0.55 13	0. 55 99
8	0.56 42	0.55 59	0.56 43	0.55 80	0. 56 38
9	0.55 75	0.56 20	0.56 27	0.55 79	0. 56 30
10	0.56 34	0.56 60	0.56 32	0.56 30	0. 57 24

4.2.6 Visualisasi Data

Visualisasi data dilakukan menggunakan WordCloud yang menampilkan data-data yang digunakan terbanyak, dan Seaborn yang menampilkan jumlah sentimen. Pada visualisasi Seaborn, dapat dilihat bahwa sebagian besar sentimen adalah sentimen negatif dan hanya sedikit sentimen positif.



Gambar 14.Visualisasi Data

4.2.7 WordCloud

Pada visualisasi WordCloud, dapat dilihat bahwa kata-kata yang sering digunakan adalah “tapera”, “potong”, dan “rakyat”

Gambar 13. WordCLoud

P-ISSN : 2580-4316
E-ISSN : 2654-8054 DOI: https://doi.org/10.37817/ikraith-informatika.v9i3

5. KESIMPULAN

Penelitian ini berhasil mengimplementasikan sistem analisis sentimen untuk memetakan opini publik terhadap kebijakan TAPERA melalui platform media sosial X. Berdasarkan hasil analisis, dapat disimpulkan bahwa mayoritas opini masyarakat cenderung bersifat negatif, dengan kekhawatiran utama yang teridentifikasi melalui kata kunci seperti "potong gaji" dan "rakyat". Dalam pengujian performa algoritma, metode Machine Learning konvensional menunjukkan efektivitas yang sangat tinggi untuk dataset ini. Model Decision Tree, Support Vector Machine (SVM), dan Logistic Regression mampu mencapai tingkat akurasi maksimal sebesar 0,99. Sebaliknya, pendekatan Deep Learning menggunakan arsitektur LSTM menghasilkan akurasi yang lebih rendah, yakni sebesar 0,5624. Hal ini menunjukkan bahwa untuk klasifikasi teks bahasa Indonesia dengan skala dataset ini, model pembelajaran mesin konvensional lebih unggul dalam memberikan hasil yang stabil dan akurat.

Saran untuk pengembangan penelitian selanjutnya adalah perlunya perluasan sumber data dari platform media sosial lain agar mendapatkan perspektif publik yang lebih komprehensif. Selain itu, penggunaan teknik penyeimbangan data (data balancing) dan penyesuaian parameter (hyperparameter tuning) yang lebih mendalam pada model Deep Learning sangat disarankan untuk meningkatkan performa akurasi pada arsitektur saraf. Penelitian mendatang juga dapat mempertimbangkan penggunaan model berbasis transformer seperti IndoBERT untuk menangani kompleksitas semantik dan konteks bahasa informal Indonesia secara lebih mendalam. Temuan ini diharapkan dapat menjadi rujukan bagi pemerintah dan pemangku kepentingan dalam mengevaluasi komunikasi kebijakan publik di masa depan.

6. UCAPAN TERIMA KASIH

Penulis memanjatkan puji syukur ke hadirat Tuhan Yang Maha Esa atas rahmat-Nya sehingga penulisan artikel ilmiah ini dapat diselesaikan sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer pada Program Studi Teknik Informatika Universitas Esa Unggul. Penulis menyampaikan apresiasi dan terima kasih yang sebesar-besarnya kepada kedua orang tua, kakak-kakak, dan keluarga besar atas doa, dukungan semangat, serta dorongan yang melimpah selama proses penyusunan karya ini. Ucapan terima kasih secara khusus ditujukan kepada Bapak Muhamad Bahrul Ulum, S.Kom., M.Kom., selaku Ketua Program Studi Teknik Informatika sekaligus Dosen Pembimbing yang telah memberikan bimbingan akademik, arahan teknis, serta waktu dalam membimbing penulisan penelitian ini hingga selesai.

DAFTAR PUSTAKA

Gurkhe, D., Pal, N., & Bhatia, R. (2014). *Effective Sentiment Analysis of Social Media Datasets using Naive Bayesian Classification*. http://goo.gl/nirajp

Hutto, C. J., & Gilbert, E. (2014). *VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text*. http://sentic.net/

Palanisamy, P., Yadav, V., & Elchuri, H. (2013). *Serendio: Simple and Practical lexicon based approach to Sentiment Analysis* (Vol. 2).

Recupero, D. R., Consoli, S., Gangemi, A., Nuzzolese, A. G., & Spampinato, D. (2014). *A Semantic Web Based Core Engine to Efficiently Perform Sentiment Analysis*. http://wit.istc.cnr.it/stlab-tools/fred

Samuel, A. L. (1959). Some Studies in Machine Learning Using the Game of Checkers. *IBM Journal of Research and Development*, *3*(3), 210–229. https://doi.org/10.1147/rd.33.0210