

# Analisis Sentimen Kebijakan Tapera pada Platform Twiter Menggunakan Algoritma Naive Bayes dan Support Vector Machine

M Ilham Abdul Shaleh

Program Studi Teknik Informatika, Universitas Mataram

## 1. PENDAHULUAN

Kebijakan Tabungan Perumahan Rakyat (Tapera) merupakan program pemerintah Indonesia yang bertujuan untuk membantu masyarakat dalam memiliki rumah. Tapera adalah skema tabungan wajib untuk pekerja, baik di sektor formal maupun informal, yang nantinya dapat digunakan untuk membeli rumah atau melunasi kredit pemilikan rumah (KPR) (Kemenko Perekonomian, 2023).

Sejak diperkenalkan pada tahun 2021, kebijakan Tapera menuai pro dan kontra di kalangan masyarakat. Beberapa media nasional seperti Kompas (2023), Detik.com (2023), dan Tempo (2023) telah memberitakan kontroversi seputar kebijakan ini. Pihak yang mendukung menilai bahwa Tapera dapat membantu masyarakat berpenghasilan rendah untuk memiliki rumah. Namun, pihak yang kontra mengkhawatirkan potensi pemotongan gaji bulanan yang dianggap memberatkan (CNN Indonesia, 2023).

Dalam situasi seperti ini, analisis sentimen di media sosial, khususnya Twitter, menjadi penting untuk memahami persepsi dan opini publik secara lebih luas. Penelitian terdahulu telah menerapkan berbagai metode untuk analisis sentimen, termasuk Naive Bayes dan Support Vector Machine (SVM). Putra et al. (2021) menggunakan metode SVM untuk menganalisis sentimen masyarakat terhadap kebijakan pembatasan sosial berskala besar (PSBB) selama pandemi COVID-19 dan menemukan bahwa SVM mampu mengklasifikasikan sentimen dengan akurasi yang tinggi.

Sementara itu, Wibowo dan Yuwono (2022) menerapkan Naive Bayes dan SVM untuk menganalisis sentimen publik terkait isu lingkungan hidup di Twitter. Hasil penelitian menunjukkan bahwa kedua metode memberikan akurasi yang baik dalam mengklasifikasikan sentimen positif, negatif, dan netral. Selanjutnya, Pratama et al. (2023) menggunakan Naive Bayes dan SVM untuk menganalisis sentimen masyarakat terhadap kebijakan cukai rokok di Indonesia, dengan Naive Bayes menunjukkan akurasi yang lebih tinggi dibandingkan SVM.

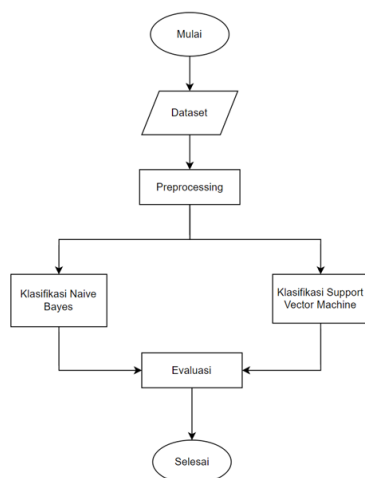
Pemilihan metode Naive Bayes dan SVM dalam penelitian ini didasarkan pada kesederhanaan algoritma, kinerja yang baik dalam klasifikasi teks, serta keberhasilan implementasi pada penelitian-penelitian sebelumnya (Feldman, 2013; Hossin and Sulaiman, 2015; Liu, 2012; Ravi and Ravi, 2015). Analisis sentimen menggunakan kedua metode ini pada kasus kebijakan Tapera penting dilakukan untuk memahami pandangan masyarakat secara komprehensif. Hal ini dapat membantu pemerintah dan pemangku kepentingan lainnya dalam mengambil tindakan yang tepat dan membuat kebijakan yang lebih baik di masa depan.

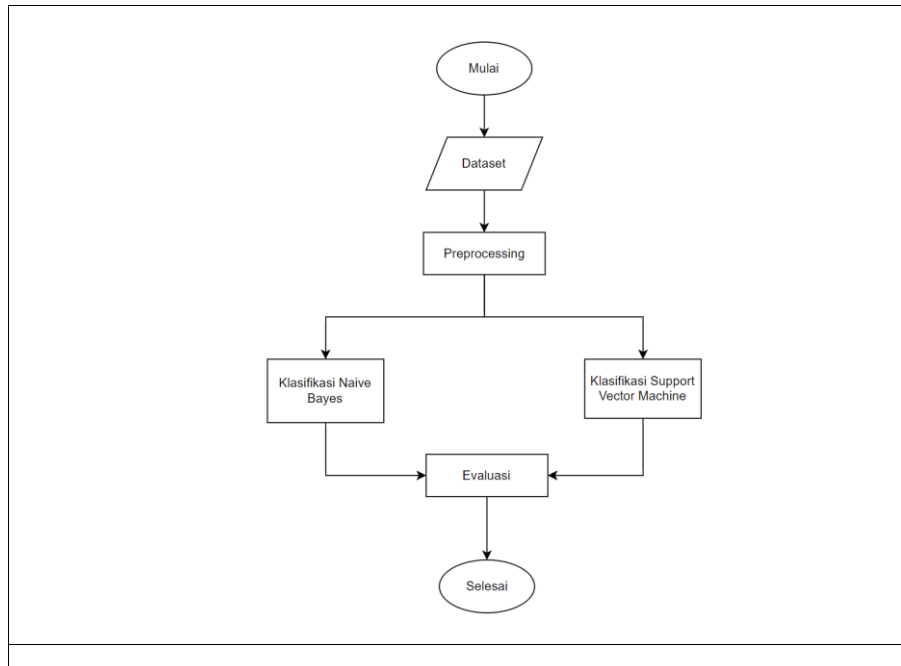
## 2. METODE

Metode yang digunakan pada penelitian ini adalah metode eksperimen. Metode ini merupakan sebuah metode yang sistematis dengan menghubungkan sebab dan akibat. Studi ini memiliki syarat yaitu; kontrol, manipulasi, dan observasi (Creswell, 2014). Eksperimen digunakan untuk mengetahui model terbaik dengan memanfaatkan data dokumen cuitan dengan dua metode algoritma SVM dan Naive Bayes untuk menganalisis sentimen terkait pembahasan Tapera (Liu, 2012).

Dalam penelitian ini, proses kontrol dilakukan dengan mengendalikan variabel-variabel yang dapat memengaruhi hasil eksperimen, seperti jenis algoritma yang digunakan dan preprocessing data. Manipulasi dilakukan dengan mengubah parameter-parameter dari algoritma SVM dan Naive Bayes untuk mendapatkan model terbaik. Observasi dilakukan dengan mengamati dan mengevaluasi performa dari kedua model yang dihasilkan.

Berikut alur penelitian yang digunakan pada penelitian ini.





**Gambar 1. Alur Penelitian**

**Memformat:** Font: 9 pt, Tebal

**Memformat:** Font: 9 pt

**Telah Diformat:** Di Tengah, Inden: Baris Pertama: 0 cm

## 2.1 Dataset

Penelitian ini menggunakan dataset yang diambil dari platform twiter menggunakan metode crawling. Data yang berhasil dikumpulkan sebanyak 2366 tweet, yang terdiri dari hasil crawling data dengan kata kunci “tapera” yang diperoleh sebanyak 1735 tweet, kata kunci “#dukungtapera” sebanyak 8 tweet, kata kunci “#iurantapera” sebanyak 59 tweet, kata kunci “#ruutapera” sebanyak 61 tweet, kata kunci “#RUUTapera” sebanyak 34 tweet, kata kunci “#TabunganPerumahanRakyat” sebanyak 95 tweet, kata kunci “#tapera” sebanyak 105 tweet, kata kunci “#Tapera” sebanyak 129 tweet, kata kunci “#tolaktapera” sebanyak 111 tweet, dan kata kunci “#uutapera” sebanyak 39 tweet.

## 2.2 Preprocessing

Tahap preprocessing bertujuan untuk mempersiapkan data agar dapat diolah dengan lebih baik oleh algoritma analisis sentimen. Dalam penelitian ini, beberapa tahapan preprocessing yang dilakukan meliputi:

### 2.2.1 Penghapusan data duplikat (remove duplicate)

Langkah ini dilakukan untuk menghilangkan data cuitan yang sama atau duplikat. Hal ini penting untuk menjaga konsistensi dan menghindari bias dalam analisis.

### 2.2.2 Pembersihan data (text cleaning)

Pada tahap ini, dilakukan penghapusan karakter-karakter yang tidak diinginkan, seperti tanda baca, URL, mention, hashtag, dan sebagainya. Tujuannya adalah untuk mempertahankan hanya teks yang relevan dengan analisis sentimen.

#### 2.2.3 Tokenisasi

Tokenisasi adalah proses memecah teks menjadi satuan-satuan yang lebih kecil, seperti kata atau kalimat. Dalam konteks analisis sentimen, tokenisasi dilakukan untuk memisahkan setiap kata dalam cuitan agar dapat diproses lebih lanjut.

#### 2.2.4 Penghapusan stopword (stopword removal)

Stopword adalah kata-kata umum yang tidak memiliki makna signifikan dalam analisis, seperti "dan," "yang," "itu," dan sebagainya. Penghapusan stopwords dilakukan untuk meningkatkan efisiensi dan akurasi analisis.

#### 2.2.5 Stemming

Stemming adalah proses untuk mendapatkan bentuk kata dasar (root word) dari sebuah kata. Misalnya, kata "belajar," "belajaran," dan "pelajaran" akan direduksi menjadi "ajar." Stemming berguna untuk menyederhanakan kata-kata dan meningkatkan akurasi analisis.

### 2.3 Naive Bayes

Dalam penelitian ini, Naive Bayes diimplementasikan untuk mengklasifikasikan sentimen cuitan terkait kebijakan Tapera (Tabungan Perumahan Rakyat) menjadi tiga kelas, yaitu positif, negatif, dan netral. Algoritma ini menghitung probabilitas suatu cuitan termasuk dalam setiap kelas berdasarkan frekuensi kemunculan kata-kata dalam data pelatihan.

### 2.4 Support Vector Machine

Dalam penelitian ini, SVM diimplementasikan dengan menggunakan kernel Radial Basis Function (RBF) untuk mengklasifikasikan sentimen cuitan terkait kebijakan Tapera menjadi tiga kelas yang sama seperti Naive Bayes, yaitu positif, negatif, dan netral. SVM berusaha menemukan hyperplane terbaik yang memisahkan kelas-kelas sentimen berdasarkan fitur-fitur yang diekstrak dari data cuitan.

### 2.5 Evaluasi

Untuk mengevaluasi performa kedua algoritma (Naive Bayes dan SVM) dalam mengklasifikasikan sentimen cuitan terkait kebijakan Tapera, digunakan metrik-metrik seperti akurasi, presisi, recall, dan F1-score.

## 3. HASIL DAN PEMBAHASAN

Berikut merupakan hasil penelitian analisis sentimen mengenai kebijakan tapera pada media sosial Twitter menggunakan algoritma naive bayes dan support vector machine (SVM). Hasil dari penelitian disajikan dalam grafik dan tabel untuk lebih mudah dipahami.

**Memformat:** Font: 11 pt, Tidak Tebal

**Memformat:** Font: 11 pt, Tidak Tebal

**Telah Diformat:** Normal, Diratakan dua sisi, Inden: Baris Pertama: 0,63 cm, Tanpa poin atau penomoran

**Memformat:** Font: 11 pt, Tidak Tebal

**Memformat:** Font: 11 pt, Tidak Tebal

### 3.1 Hasil Crawling

Pengumpulan dataset dilakukan menggunakan tools Tweet Harvest dengan memanfaatkan kode autentikasi twitter dan memasukkan kata kunci yang diinginkan, dengan total dataset yang berhasil diperoleh sebanyak 2366 tweet.

	conversation_id_str	favorite_count	id_str	quote_count	reply_count	retweet_count	user_id_str
count	2.366000e+03	2366.000000	2.366000e+03	2366.000000	2366.000000	2366.000000	2.366000e+03
mean	1.736349e+18	453.196534	1.736357e+18	31.257396	30.989856	134.636940	5.988125e+17
std	2.626369e+17	2787.702157	2.626388e+17	276.456795	194.299803	811.543148	6.958610e+17
min	2.171015e+17	0.000000	2.171015e+17	0.000000	0.000000	0.000000	6.689892e+06
25%	1.797822e+18	0.000000	1.797823e+18	0.000000	0.000000	0.000000	1.357955e+08
50%	1.798185e+18	3.000000	1.798199e+18	0.000000	0.000000	1.000000	2.892770e+09
75%	1.798685e+18	16.000000	1.798686e+18	1.000000	2.000000	5.000000	1.322361e+18
max	1.801432e+18	54778.000000	1.801432e+18	8685.000000	3991.000000	13449.000000	1.793151e+18

Gambar 2. Statistik Deskriptif Dataset

### 3.2 Preprocessing

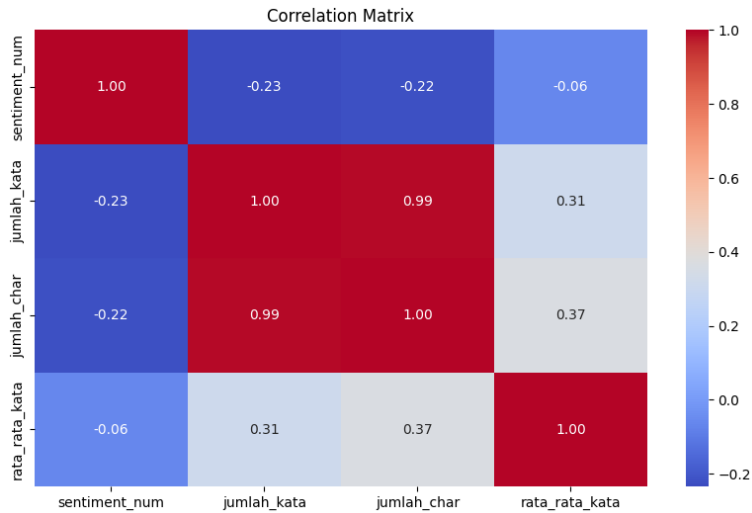
Proses cleaning dilakukan untuk membersihkan teks dari karakter-karakter yang tidak diperlukan seperti URL, mention, hashtag, simbol, angka, dan lainnya. Langkah-langkahnya meliputi penggunaan ekspresi regular untuk menghapus URL, mention, dan hashtag, serta karakter khusus lainnya. Selanjutnya, teks diubah menjadi lowercase untuk menyamakan format huruf, diikuti dengan penghapusan stopwords yang tidak relevan menggunakan daftar stopwords dari NLTK. Kemudian, kata-kata dalam teks diubah menjadi bentuk dasarnya menggunakan lemmatisasi dengan WordNetLemmatizer dan stemming dengan PorterStemmer dari NLTK.

	tweet_clean	full_text
255	yang mulia pinokio kalian itu lho kok percaya ...	@democrazymedia Yang Mulia Pinokio: Kalian itu...
359	yang mulia pinokio kalian itu lho kok percaya ...	@democrazymedia Yang Mulia Pinokio: Kalian itu...
521	investasi bodong tapera	Investasi Bodong Tapera #TolakTapera https://t...
695	tapera itu sebuah bidang inovasi pembiayaan ta...	Tapera itu sebuah bidang inovasi pembiayaan ta...

Gambar 3. Contoh Hasil Cleaning Dataset

### 3.3 Visualisasi Data

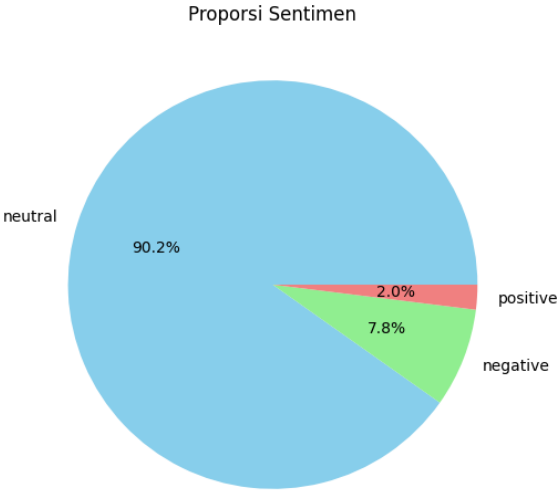
3.3.1 Correlation Matrix



Memformat: Font: Tebal

Gambar 4. Heatmap Correlation Matrix

3.3.2 Perbandingan Sentimen



Memformat: Font: 9 pt

Memformat: Font: 9 pt

Memformat: Font: 9 pt

Telah Diformat: Normal, Di Tengah, Inden: Kiri: 2,14 cm, Tanpa poin atau penomoran

Memformat: Font: Tebal

Gambar 5. Proporsi Sentimen

Telah Diformat: Kiri, Inden: Kiri: 6,58 cm, Tanpa poin atau penomoran

Memformat: Font: Tebal

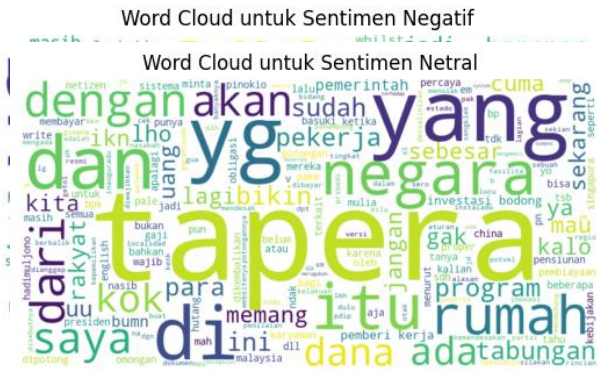
Telah Diformat: Normal, Tanpa poin atau penomoran

3.3.3 Word Cloud



Gambar 6. Word Cloud Sentimen Positif

Gambar 7. Word Cloud Sentimen Negatif



Gambar 8. Word Cloud Sentimen Netral

Memformat: Font: Tebal

Memformat: Font: 9 pt

Memformat: Font: 9 pt

Memformat: Font: 9 pt

Telah Diformat: Normal, Inden: Kiri: 5,95 cm, Baris Pertama: 0,4 cm, Tanpa poin atau penomoran

Memformat: Font: Tebal

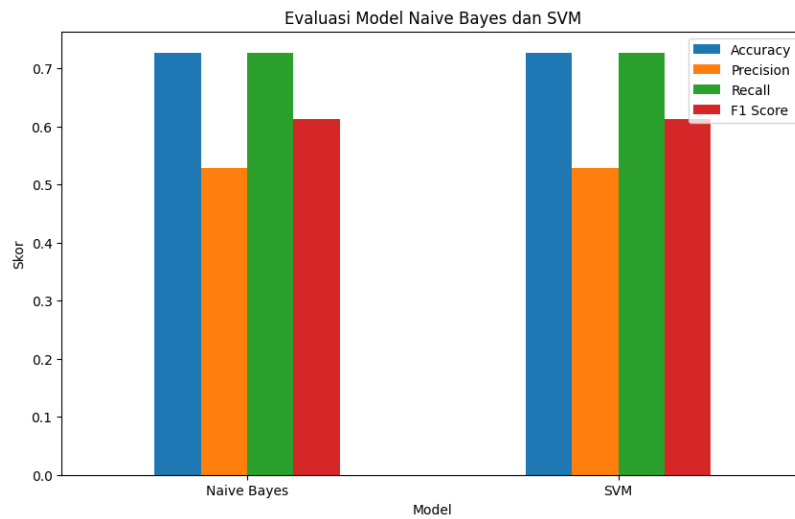
Telah Diformat: Normal, Tanpa poin atau penomoran

Memformat: Font: Tebal

Telah Diformat: Inden: Kiri: 1,5 cm, Tanpa poin atau penomoran

### 3.4 Hasil Modeling

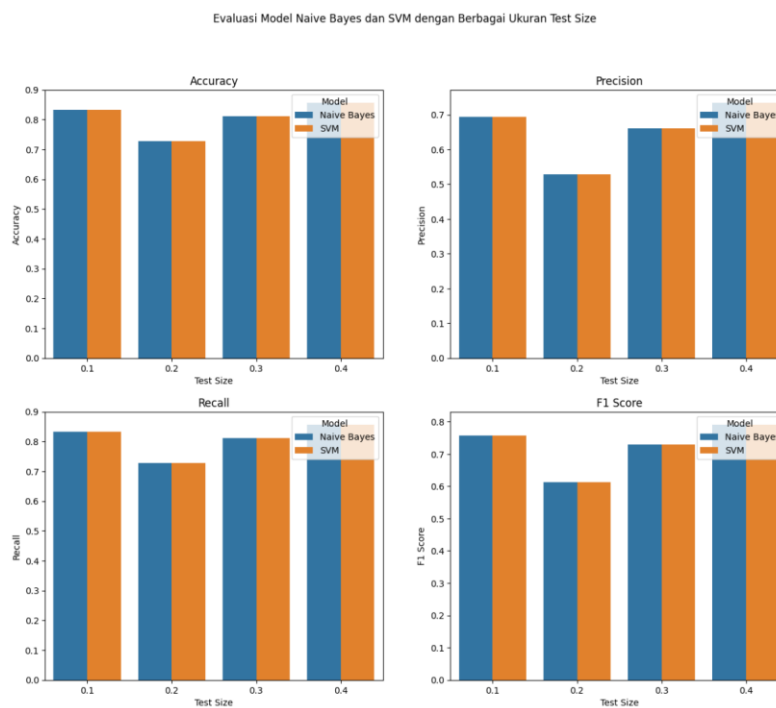
**Memformat:** Font: Tidak Tebal



**Gambar 9.** Hasil Evaluasi Model

### 3.5 Perbandingan Berbagai Ukuran Set Pengujian

**Telah Diformat:** Inden: Kiri: 5,08 cm, Baris Pertama: 1,27 cm, Tanpa poin atau penomoran



**Gambar 9.** Hasil Perbandingan Evaluasi Model Berbagai Set Pengujian

**Memformat:** Font: 9 pt

**Telah Diformat:** Normal, Inden: Kiri: 2,77 cm, Baris Pertama: 1,04 cm, Tanpa poin atau penomoran



#### 4. PENUTUP

Berdasarkan data hasil evaluasi performa algoritma Naive Bayes dan Support Vector Machine (SVM) dalam mengklasifikasikan sentimen cuitan terkait kebijakan Tapera, dapat dilihat bahwa kedua algoritma menunjukkan kinerja yang cukup baik secara keseluruhan.

Pada ukuran test size 0.4, baik Naive Bayes maupun SVM memiliki akurasi, presisi, recall, dan F1-score yang sama, yaitu 0.857143, 0.734694, 0.857143, dan 0.791209 secara berturut-turut. Ini menunjukkan bahwa kedua algoritma memiliki performa yang seimbang pada ukuran test size yang lebih besar.

Sementara itu, pada ukuran test size yang lebih kecil, seperti 0.3, 0.2, dan 0.1, kinerja kedua algoritma juga masih cukup baik, meskipun terdapat sedikit penurunan pada nilai-nilai metrik tertentu. Namun, pola yang terlihat adalah semakin kecil ukuran test size, semakin rendah nilai akurasi, presisi, recall, dan F1-score yang diperoleh.

Secara keseluruhan, dari data ini dapat disimpulkan bahwa baik Naive Bayes maupun SVM memiliki kinerja yang cukup baik dalam mengklasifikasikan sentimen cuitan terkait kebijakan Tapera, terutama pada ukuran test size yang lebih besar. Pemilihan algoritma yang tepat dapat disesuaikan dengan kebutuhan dan karakteristik data yang dimiliki.

**Telah Diformat:** Daftar Paragraf, Inden: Kiri: 1,5 cm

**Memformat:** Font: Tidak Tebal

**Telah Diformat:** Diratakan dua sisi, Inden: Baris Pertama: 0,75 cm

**DAFTAR PUSTAKA**

- Feldman, R. (2013). Techniques and applications for sentiment analysis. Communications of the ACM, 56(4), 82-89.
- Hossin, M., & Sulaiman, M. N. (2015). A review on evaluation metrics for data classification evaluations. International Journal of Data Mining & Knowledge Management Process, 5(2), 1-11.
- Liu, B. (2012). Sentiment analysis and opinion mining. Synthesis lectures on human language technologies, 5(1), 1-167.
- Pratama, A., Putri, D. A., & Nurhuda, F. (2023). Analisis Sentimen Masyarakat Terhadap Kebijakan Cukai Rokok Menggunakan Naive Bayes dan Support Vector Machine. Jurnal Ilmiah Teknologi Informasi, 21(1), 1-8.
- Putra, R. P., Pratama, A., & Sari, Y. A. (2021). Analisis Sentimen Masyarakat Terhadap Kebijakan PSBB Menggunakan Metode Support Vector Machine. Jurnal Informatika dan Rekayasa Komputer, 5(2), 115-122.
- Ravi, K., & Ravi, V. (2015). A survey on opinion mining and sentiment analysis: Tasks, approaches and applications. Knowledge-Based Systems, 89, 14-46.
- Wibowo, A., & Yuwono, F. (2022). Analisis Sentimen Publik Terhadap Isu Lingkungan Hidup di Media Sosial Twitter Menggunakan Metode Naive Bayes dan Support Vector Machine. Jurnal Sains dan Teknologi, 12(2), 123-135.

**Memformat:** Font: 12 pt**Telah Diformat:** Normal, Diratakan dua sisi, Inden: Baris Pertama: 0,75 cm, Tanpa poin atau penomoran**Memformat:** Font: Tebal

**LINK TERKAIT YANG DIPERLUKAN**

Link Youtube : <https://youtu.be/GdGCsC87xJo>

Link PPT : <https://sprl.in/projectakhirilhamppt>

Repository Github : <https://github.com/mhmadilham18/Analisis-Sentimen-Tapera>