

# Analisis Sentimen Kebijakan Tabungan Perumahan Rakyat (Tapera) Menggunakan *Support Vector Machine*

Aziz Musthafa<sup>1</sup>, Triana Harmini<sup>2</sup>, Abid Rafiq<sup>3</sup>, Nurhana Marantika<sup>4</sup>

Teknik Informatika

Universitas Darussalam Gontor

Ponorogo, Indonesia

e-mail: <sup>1</sup>aziz@unida.gontor.ac.id, <sup>2</sup>triana@unida.gontor.ac.id, <sup>3</sup>abidrafiq@mhs.unida.gontor.ac.id,

<sup>4</sup>hana@unida.gontor.ac.id

Correspondence : e-mail: abidrafiq@mhs.unida.gontor.ac.id

Diajukan: 21 Agustus 2024; Direvisi: 24 Agustus 2024; Diterima: 25 Agustus 2024

## Abstrak

Tabungan Perumahan Rakyat (Tapera) merupakan kebijakan yang diadakan oleh pemerintah untuk membantu masyarakat terutama Masyarakat Berpenghasilan Rendah (MBR) dalam memenuhi kebutuhan tempat tinggal yang layak. Kebijakan ini menuai banyak pro dan kontra bagi masyarakat termasuk di media-media sosial seperti Twitter karena dipandang bahwa Tapera hanya akan menambah beban finansial mereka. Penelitian ini bertujuan untuk melakukan analisis sentimen klasifikasi pandangan masyarakat terhadap kebijakan Tapera pada media sosial Twitter. Dataset penelitian diambil melalui Twitter dengan teknik crawling yang kemudian diklasifikasikan dalam sentimen positif, negatif, dan netral. Metode yang digunakan adalah Support Vector Machine (SVM) dan hasil penelitian dievaluasi menggunakan Cross Validation sebagai validator. Hasil penelitian menunjukkan bahwa model SVM penelitian ini menghasilkan akurasi sebanyak 87% dengan nilai precision sebesar 87%, recall sebesar 88%, dan f1-score sebesar 88%.

**Kata kunci:** Analisis Sentimen, Tapera, Support Vector Machine.

## Abstract

Public Housing Savings (Tapera) is a policy implemented by the government to help the community, especially Low Income Communities (MBR), in meeting their needs for adequate housing. This policy has many pros and cons for the public, including on social media such as Twitter because it is believed that Tapera will only increase their financial burden. This research aims to carry out sentiment analysis to classify people's views on the Tapera policy on Twitter social media. The research dataset is taken via Twitter using a crawling technique which is then classified into positive, negative and neutral sentiment. The method used is Support Vector Machine (SVM) and the research results are evaluated using Cross Validation as a validator. The research results show that the SVM model of this research produces an accuracy of 87% with a precision value of 87%, recall of 88%, and f1-score of 88%.

**Keywords:** Sentiment Analysis, Tapera, Support Vector Machine.

## 1. Pendahuluan

Pemerintah memiliki tanggung jawab untuk menyediakan perumahan yang terjangkau bagi masyarakat. Namun, tantangan dalam memenuhi permintaan perumahan yang terjangkau masih menjadi masalah di Indonesia. Di Indonesia, terdapat beberapa kebijakan subsidi perumahan yang diterapkan oleh pemerintah, seperti KPR FLPP, Subsidi Bunga Kredit Perumahan, dan Roadmap Sistem Pembiayaan Perumahan Indonesia.[1] Berdasarkan data BPS, persentase rumah tangga yang memiliki akses terhadap rumah layak huni selalu mengalami peningkatan sejak tahun 2019 hingga tahun 2021, namun peningkatan tiap tahunnya cenderung tidak signifikan. Provinsi Papua memiliki persentase terendah 40,81 % sementara Provinsi Daerah Istimewa Yogyakarta menempati posisi tertinggi 97,12%. Berdasarkan data BPS menunjukkan bahwa tingkat persentase rumah tangga yang memiliki akses terhadap rumah layak huni daerah perkotaan selalu lebih tinggi dibandingkan daerah perdesaan. Sehingga masih terdapat ketidakmerataan persentase rumah layak huni di setiap di Indonesia.[2]

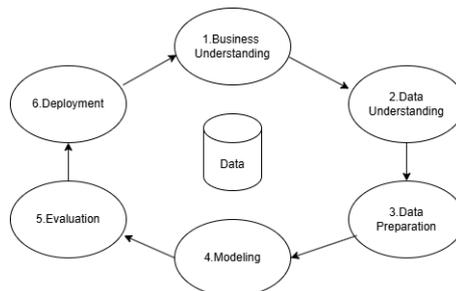
Penerapan Tapera di Indonesia masih menuai begitu banyak pro dan kontra. Pihak yang mendukung kebijakan Tapera berpendapat bahwa Tapera dapat menjadi solusi pemerataan sehingga semua masyarakat dapat memiliki tempat tinggal permanen yang murah dan layak huni. Namun, bagi pihak yang kontra menganggap bahwa program Tapera hanya akan menambah beban pemberi kerja dan mengurangi upah yang seharusnya diterima oleh para pekerja. [3]

Menurut APJII, sekitar 51,5% pengguna internet di Indonesia menggunakan media sosial setiap hari. Salah satu media sosial yang sering digunakan di Indonesia adalah *Twitter*. [4] Analisis sentimen merupakan proses analisis teks digital untuk memahami pendapat sikap dan emosi seseorang ketika mengungkapkan pendapat dalam teks seperti pada review, berita, serta komentar media sosial. *Support Vector Machine (SVM)* merupakan algoritma yang populer digunakan dalam analisis sentimen karena dapat mengklasifikasi data dengan jumlah yang besar dengan hasil yang akurat. [5]

Penelitian terdahulu dalam analisis sentimen pada *platform Twitter* telah banyak dilakukan [6]. penelitian [7] menggunakan algoritma SVM untuk menganalisa vaksin *booster*. Beberapa penelitian menggunakan SVM untuk menganalisa sentimen produk kartu prakerja di *Twitter*. [8] Adapun analisis sentimen terkait kebijakan tapera sudah ada namun masih sedikit. [9] Oleh karena itu, penelitian ini dibuat untuk mengisi celah pengetahuan terkait analisis sentimen terhadap kebijakan Tapera.

Melalui hasil penelitian ini, diharapkan dapat berkontribusi dalam pengembangan model analisis sentimen terhadap *platform twitter*, kontribusi lain dari penelitian ini diharapkan agar literasi terkait analisis sentimen khususnya Tapera dapat meningkat. hasil dari penelitian ini juga dapat membantu memahami sentimen masyarakat terhadap kebijakan Tapera.

## 2. Metode Penelitian



Gambar 1. Model CRISP-DM.

Penelitian ini menggunakan *Cross Industry Standard Process for Data Mining (CRISP-DM)*. CRISP-DM merupakan kerangka kerja umum yang digunakan dalam pemrosesan data untuk membantu pengguna dalam memahami dan mengelola proyek data mining secara sistematis. [10] Model ini memiliki enam tahapan : *Business Understanding*, *Data Understanding*, *Data preparation*, *Modeling*, *Evaluation*, dan yang terakhir *Deployment*. Pada Gambar 1 menunjukkan Tahapan CRISP – DM.

### 2.1. Business Understanding

Tujuan utama tahap ini adalah untuk menganalisis pandangan Masyarakat terhadap kebijakan Tapera berdasarkan media sosial *Twitter* dan memberikan wawasan yang berguna bagi pembuat kebijakan dan otoritas lingkungan untuk memahami persepsi publik.

### 2.2. Data Understanding

Pada Tahapan *Data Understanding*, Data dikumpulkan dengan menggunakan *crawling* pada media sosial *Twitter* dengan *python*. Data yang digunakan dalam penelitian ini adalah data dengan *key search* “tapera lang:id:”. Selanjutnya, data yang duplikat, atau tidak berhubungan dengan Tapera dihapus, dan data yang tersisa dilabel yang terbagi menjadi positif, negatif, dan netral kemudian divalidasi oleh ahli Bahasa Indonesia agar meminimalisir kesalahan dalam pelabelan data.

### 2.3. Data Preparation

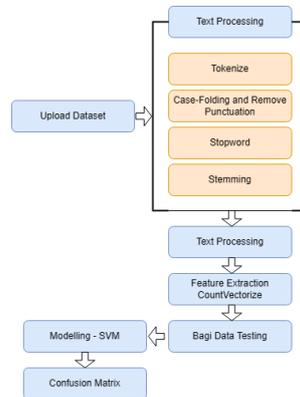
Pada tahap *Data Preparation*, Data *tweet* yang telah dikumpulkan akan melalui beberapa pemrosesan teks (*Preprocessing Data*) yang terdiri dari Normalisasi, *Stopwords*, *Tokenize*, dan *Stemming*. Data yang sudah dibersihkan kemudian diubah ke dalam bentuk vektor menggunakan fitur extraction *CountVectorizer* dan melakukan pembobotan dengan menggunakan Tf-Idf. Tf-Idf dihitung menggunakan persamaan sebagai berikut :

$$(1)$$

Dengan  $W_x$ , adalah bobot istilah (ty) terhadap dokumen (dx). Sedangkan  $tf_x$ , adalah jumlah kemunculan istilah (ty) dalam dokumen (dx). N adalah jumlah semua dokumen yang ada dalam dataset dan dfx adalah jumlah dokumen yang mengandung istilah (ty), minimal ada satu kata yaitu istilah (ty).[11]

Kemudian data tersebut dibagi menjadi dua bagian untuk keperluan *training* dan *testing* dengan perbandingan 80:20., dan dilakukan pemodelan menggunakan metode SVM dan perhitungan untuk melihat akurasi model yang telah dibuat.

**2.4. Modelling**



Gambar 2. Model algoritma SVM

Pada tahapan modelling, algoritma yang digunakan adalah *Support Vector Machine (SVM)* karena *Support Vector Machine* berusaha menemukan *Hyperplane* dengan memaksimalkan jarak antar kelas. Dengan cara ini, *Support Vector Machine* dapat menjamin kemampuan generalisasi yang tinggi untuk data-data yang akan datang. *Support Vector Machine* menemukan *Hyperplane* dengan memaksimalkan jarak antar kelas (margin). *Support Vector Machine* hanya menemukan satu *Hyperplane* yang posisinya tepat di tengah-tengah antar dua kelas. SVM memiliki persamaan sebagai berikut :

$$f(x) = \text{sign}(w*x + b) \tag{2}$$

dimana f(x) adalah fungsi hipotesis yang menentukan kelas mana yang dipilih, w adalah vektor bobot yang ditentukan selama proses pelatihan, x adalah vector fitur input, dan b adalah bias.

**2.5. Evaluation**

Untuk hasil dari evaluasi, dapat menggunakan data uji (*testing data*) yang terpisah untuk mengevaluasi kinerja model *Support Vector Machine (SVM)*. Peneliti menggunakan *Cross Validation* sebagai validator..

**2.6. Deployment**

Tahap terakhir dalam model CRISP-DM adalah tahap *Deployment* dimana model *Support Vector Machine (SVM)* yang telah dibangun dan dievaluasi diimplementasikan ke dalam lingkungan operasional.

**3. Hasil dan Pembahasan**

**3.1. Pengumpulan data**

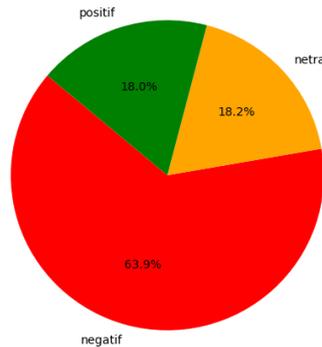
Proses pengumpulan data dilakukan menggunakan *Tweet-harvest* untuk crawling data dari Twitter. Kata kunci yang digunakan adalah “tapera lang:id” menghasilkan sebanyak 5011 data mentah dari tanggal 21 Juni 2024 sampai 23 Juli 2024. Data yang berhasil dikumpulkan kemudian dibersihkan dengan menghapus duplikat, dan membuang data yang tidak berguna untuk penelitian secara manual dengan jumlah akhir sebanyak 2936 data. Kemudian, proses labeling divalidasi oleh dosen pakar dari Ilmu Komunikasi agar tidak terjadi kesalahan dalam proses label. Contoh hasil labeling data dapat dilihat pada Tabel 3.1.

Tabel 1. Contoh Hasil *Crawling* Dan *Labeling*

Created at	Full Text	Username	Label
Tue Jul 23 15:42:07 +0000 2024	@detikfinance Syaratnya korupsi banyak banyak gadaikan kalimantan 190 thm dan ambil pajak rakyat lewat tapera dll	kancilakal71566	negatif

Terlihat dari Gambar 3.1 bahwa komentar negatif memiliki nilai tertinggi berjumlah 1874 (63.9%) data, diikuti oleh komentar netral dengan jumlah 534 (18.2%), dan komentar positif berjumlah 527 (18.0%). Menunjukkan jumlah komentar negatif lebih mendominasi dari komentar-komentar lainnya

Perbandingan banyaknya data komentar positif, netral, dan negatif



Gambar 2. Pie-Chart Dataset.

### 3.2. Preprocessing

Pada tahap *preprocessing*, data yang sudah dikumpulkan akan diubah dan diproses menjadi data bersih sebelum menuju tahap pemrosesan menggunakan model SVM. Contoh hasil dari *preprocessing* data dapat dilihat pada Tabel.3.2.

Tabel 3.2. Contoh Hasil Preprocessing

Full Text	Case-folding	Tokenizing	Normalize	Stopwords	Stemming
@detikfinance Syaratnya korupsi banyak banyak gadaikan kalimantan 190 thm dan ambil pajak rakyat lewat tapera dll	syaratnya korupsi banyak banyak gadaikan kalimantan thm dan ambil pajak rakyat lewat tapera dll	['syaratnya', 'korupsi', 'banyak', 'banyak', 'gadaikan', 'kalimantan', 'thm', 'dan', 'ambil', 'pajak', 'rakyat', 'lewat', 'tapera', 'dll']	syaratnya korupsi banyak banyak gadaikan kalimantan thm dan ambil pajak rakyat lewat tapera dan lain-lain	syaratnya korupsi banyak banyak gadaikan kalimantan thm dan ambil pajak rakyat lewat tapera dan lain-lain	syarat korupsi gadai kalimantan thm ambil pajak rakyat tapera lain

#### 3.2.1. Case-folding

Tahap *Case-folding* mengubah semua huruf disetiap kalimat pada data menjadi huruf kecil. Contoh dari *case-folding* seperti “Kalimantan” setelah diproses menjadi “kalimantan”.

#### 3.2.2. Normalisasi

Tahap normalisasi mengubah semua kata-kata dalam komentar menjadi sesuai dengan aturan bahasa Indonesia. Komentar seperti “dll” diubah menjadi “dan lain-lain”

#### 3.2.3. Stopwords

Tahap *stopwords* removal menghilangkan kata-kata yang tidak penting atau tidak memiliki arti. Contoh dari kata-kata *stopwords* seperti “eh”, “loh”, “nah”, dan “oh”.

#### 3.2.4. Stemming

Tahap *stemming* mengubah kata-kata yang telah diproses menjadi kata-kata dasar. Contoh dari *stemming* seperti kata “membayar” menjadi “bayar”.

### 3.3. Klasifikasi Support Vector Machine

Data yang telah dikumpulkan dan diproses dalam tahap preprocessing kemudian dilatih dengan model SVM kernel Linear untuk membagi data menjadi 3 kelas yaitu positif, netral, dan negatif dengan data uji sebanyak 1125.

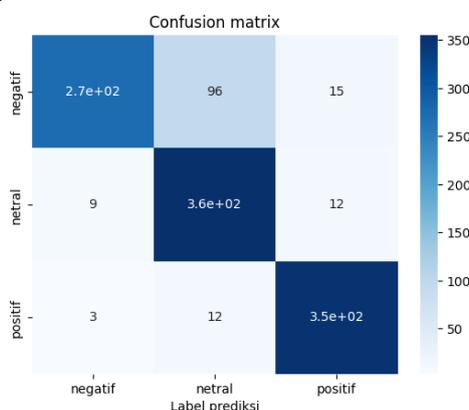
### 3.4. Evaluasi

Model *Support Vector Machine* yang sudah dibuat kemudian dievaluasi dengan *confusion matrix*. Tabel 3.4 menunjukkan bahwa nilai *precision* komentar negatif mencapai sebesar 0.96, *recall* sebesar 0.71, *f1-score* sebesar 0.82. Sedangkan nilai *precision* komentar netral mencapai 0.77, *recall* sebesar 0.94, *f1-score* sebesar 0.85. Adapun nilai *precision* komentar positif mencapai 0.93, *recall* sebesar 0.96, dan *f1-score* sebesar 0.94. Nilai akurasi keseluruhan yang didapat adalah 0.87.

Tabel 3.3. Nilai akurasi confusion matrix

	Precision	Recall	f1-score
<b>Negatif</b>	0.96	0.71	0.82
<b>Netral</b>	0.77	0.94	0.85
<b>Positif</b>	0.93	0.96	0.94
<b>Accuracy</b>		0.87	

Pada Gambar 2, model SVM yang dibuat berhasil mengklasifikasikan nilai negatif dengan *True Negative (TN)* sebesar 270, *False Neutral (FNT)* sebesar 94, dan *False Positive (FP)* sebesar 15. Kemudian *False Negative (FN)* sebesar 9, *True Neutral (TNT)* sebesar 360, dan *False Positive (FP)* sebesar 12. Adapun nilai *False Negative (FN)* sebesar 3, *False Neutral (FNT)* sebesar 12, dan *True Positive (TP)* sebesar 350. Model SVM yang dibuat sudah berhasil mengklasifikasikan sebagian besar dataset sesuai label masing-masing dengan benar.



Gambar 2. Confusion Matrix

Banyaknya frekuensi kata yang muncul digambarkan melalui *wordcloud* pada Gambar 3, dimana kata “tapera”, “rakyat”, “asuransi”, dan “pajak”. memiliki frekuensi kemunculan yang tinggi. Gambar 4. menunjukkan 10 kata yang paling banyak muncul dengan kata “tapera” menduduki peringkat paling tinggi dengan frekuensi sebanyak 1938 kali, dan “kendara” yang menduduki paling rendah dari kesepuluh kata dibawah dengan frekuensi sebanyak 155 kali. Wordcloud ini secara keseluruhan menunjukkan bahwa kebijakan Tapera direspon sangat negatif dengan ketidakpuasan masyarakat, terutama terkait dengan dampak pada finansial pribadi, serta kekhawatiran dari keadilan dan kejelasan dalam pelaksanaannya. Pihak pemangku kebijakan dapat menanggapi respon ini dengan mempertimbangkan dan memperbaiki sistematisa penerapan Tapera, lalu mengawasi pelaksanaannya dengan ketat agar masyarakat tidak merasa terbebani dan memberikan sosialisasi yang lebih jelas agar tidak menimbulkan keraguan.



- 
- [7] H. Saputra, “Analisis Sentimen Pada Vaksin Booster Menggunakan Algoritma *Support Vector Machine* Multiclass Di Twitter,” *J. Teknol. Pint.*, vol. 3, no. 10, pp. 1–26, 2023, [Online]. Available: <http://teknologipintar.org/index.php/teknologipintar/article/view/506>
- [8] S. Styawati, N. Hendrastuty, and A. R. Isnain, “Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode *Support Vector Machine*,” *J. Inform. J. Pengemb. IT*, vol. 6, no. 3, pp. 150–155, 2021, doi: 10.30591/jpit.v6i3.2870.
- [9] E. D. Harahap and R. Kurniawan, “Analisis Sentimen Komentar Terhadap Kebijakan Pemerintah Mengenai Tabungan Perumahan Rakyat ( TAPERA ) Pada Aplikasi X Menggunakan Metode Naïve Bayes,” *J. Tek. Inform. Unika ST. Thomas*, vol. 09, no. 1, 2024.
- [10] C. Schröer, F. Kruse, and J. M. Gómez, “A systematic literature review on applying CRISP-DM process model,” *Procedia Comput. Sci.*, vol. 181, no. 2019, pp. 526–534, 2021, doi: 10.1016/j.procs.2021.01.199.
- [11] E. Sera, Hazriani, Mirfan, and Yuyun, “Analisis Sentimen Ulasan Produk di E-Commerce Bukalapak Menggunakan Natural Language Processing,” *Pros. SISFOTEK*, pp. 237–243, 2023, [Online]. Available: <http://www.seminar.iaii.or.id/index.php/SISFOTEK/article/view/406%0Ahttp://www.seminar.iaii.or.id/index.php/SISFOTEK/article/download/406/338>