

Analisis Sentimen Masyarakat Terhadap Tapera Pada Media Sosial X Menggunakan Metode *K-Nearest Neighbor*

Muhammad Iksan Alfandi¹⁾, Pitrasacha Adytia²⁾, dan Wahyuni³⁾

¹Program Studi Sistem Informasi, Sekolah Tinggi Manajemen Informatika dan Komputer Widya Cipta Dharma

¹Jl. M. Yamin No.25, Samarinda, 75123

E-mail: iksanalfn@gmail.com¹⁾

ABSTRAK

Program Tapera (Tabungan Perumahan Rakyat) yang ingin diluncurkan untuk membantu masyarakat untuk memiliki rumah yang layak huni adalah tujuan dari pemerintah Indonesia. Penggunaan dana, persyaratan kontribusi, dan isu-isu lainnya bukannya tidak ada perdebatan mengenai program ini. Dengan banyaknya pengguna dari berbagai latar belakang sosial, Media Sosial X menghasilkan berbagai macam komentar dan sudut pandang terhadap setiap berita yang sedang diberitakan. Ada orang yang pro, netral, dan anti terhadap Program Tapera. Tujuan dari penelitian ini adalah untuk melakukan analisis sentimen pengguna Media Sosial X terhadap topik-topik yang berkaitan dengan kebijakan pemerintah tentang Tapera dengan menggunakan algoritma K-NN (K-Nearest Neighbor). Penelitian ini menggunakan hingga 958 tweet dalam bahasa Indonesia yang dikumpulkan dari tanggal 19 Juli 2024 hingga 23 Juli 2024. Nilai sentimen akan dikategorikan ke dalam tiga kelas positif, netral, dan negatif setelah kata-kata tersebut diberi bobot menggunakan TF-IDF. Tiga nilai - $K = 3$, $K = 5$, dan $K = 7$ diuji coba. Setelah dilakukan pengujian dengan nilai K , $K=3$ dengan akurasi 62,75% menghasilkan nilai akurasi tertinggi. Ditemukan bahwa persepsi masyarakat terhadap Tapera cenderung kurang baik.

Kata Kunci: Sentimen, X, Tapera, *K-Nearest Neighbor*

ABSTRACT

The Tapera (Public Housing Saving) program, which aims to help people own a decent home, is a goal of the Indonesian government. The use of funds, contribution requirements, and other issues are not without debate regarding this program. With a large number of users from various social backgrounds, Social Media X generates a wide variety of comments and viewpoints on any news that is being reported. There are people who are pro, neutral, and anti the Tapera Program. The purpose of this study is to conduct sentiment analysis of Social Media X users on topics related to government policies on Tapera using the K-NN (K-Nearest Neighbor) algorithm. This study uses up to 958 tweets in Indonesian language collected from July 19, 2024 to July 23, 2024. The sentiment values will be categorized into three classes of positive, neutral, and negative after the words are weighted using TF-IDF. Three values - $K=3$, $K=5$, and $K=7$ were tested. After testing with K values, $K=3$ with 62.75% accuracy resulted in the highest accuracy value. It was found that the public perception of Tapera tends to be unfavorable.

Keywords: Sentiment, X, Tapera, *K-Nearest Neighbor*

1. PENDAHULUAN

Permasalahan krisis perumahan di Indonesia telah menjadi fokus utama dalam beberapa dekade terakhir. Meskipun berbagai inisiatif telah diluncurkan untuk mengatasi kekurangan perumahan yang signifikan, tantangan dalam menyediakan perumahan yang terjangkau dan layak tetap mempengaruhi jutaan orang di seluruh negeri.

Tapera, yang diperkenalkan melalui Undang-Undang Nomor 4 Tahun 2016, memiliki tujuan utama untuk memberikan perlindungan sosial dan memfasilitasi akses terhadap perumahan yang terjangkau bagi seluruh lapisan masyarakat. Program ini didesain sebagai bentuk integrasi antara upaya penyediaan perumahan dan pengembangan

keuangan inklusif, dengan harapan dapat mengurangi disparitas perumahan yang luas di berbagai wilayah Indonesia.

Di bawah kerangka hukum Tapera, para pekerja formal diwajibkan untuk menyisihkan sebagian kecil dari penghasilan mereka setiap bulan ke dalam tabungan perumahan, yang nantinya dapat digunakan untuk membiayai pembelian, perbaikan, atau pemeliharaan perumahan mereka.

Namun, meskipun potensi manfaatnya yang signifikan, implementasi Tapera tidak lepas dari tantangan dan kritik. Perlu adanya strategi yang matang dalam pengelolaan dana, peningkatan kesadaran masyarakat

terhadap pentingnya tabungan perumahan, serta pemantauan yang ketat terhadap keberlanjutan dan transparansi program.

Hasil survei terbaru dari Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) menunjukkan bahwa tingkat penetrasi internet di Indonesia meningkat sebesar 77,02%, dengan jumlah pengguna internet mencapai 210 juta dari 272 juta penduduk Indonesia yang telah terkoneksi internet pada tahun 2021. Tingkat penetrasi internet dan kontribusinya hampir sama di seluruh gender. Hal ini menunjukkan bahwa tidak ada perbedaan, karena banyak orang sekarang menganggap memiliki akses ke internet sebagai sebuah kebutuhan. Kelompok usia yang paling banyak menikmati *cybersurfing* adalah 13 hingga 18 tahun. Kelompok usia 19-34 tahun berada di urutan berikutnya, dengan perbedaan yang cukup kecil. Kelompok usia 35 hingga 54 tahun berada di urutan berikutnya.

Menurut diskusi survei, mayoritas dari mereka adalah pengguna internet yang mengakses media sosial. Lima platform media sosial teratas dalam hal basis pengguna adalah *X*, *Facebook*, *Instagram*, *Youtube*, dan *Tiktok*.

Karena Media Sosial *X* sangat mudah digunakan, orang Indonesia menggunakannya secara luas. Yang perlu didaftarkan oleh pengguna hanyalah alamat email dan nomor ponsel. Fitur tweet dari *X* adalah salah satu fitur yang paling banyak digunakan.

Pengguna dapat menulis ide, opini, dan pemikiran mereka dengan memanfaatkan fungsi ini. Tidak diragukan lagi, *tweet* pengguna platform ini dapat dianalisis untuk menghasilkan data yang mendalam.

Setiap orang memiliki pemikiran dan sudut pandang yang unik. Sudut pandang ini sangat penting dan merupakan salah satu faktor yang terutama membentuk perilaku manusia. Analisis sentimen, terkadang dikenal sebagai penggalan opini, adalah cabang penelitian yang meneliti bagaimana perasaan masyarakat umum tentang berbagai topik, termasuk peristiwa yang terjadi di masyarakat secara keseluruhan serta isu, masalah, dan opini tentang tokoh, organisasi, layanan, atau produk tertentu. (Latif, 2021).

Dalam kehidupan sehari-hari, analisis sentimen digunakan secara luas di berbagai sektor. Analisis sentimen, misalnya, digunakan dalam dunia bisnis untuk meramalkan harga saham, menilai permintaan konsumen, dan membuat rencana pemasaran berdasarkan sentimen di kalangan masyarakat umum. Analisis sentimen juga digunakan untuk mengetahui pendapat masyarakat umum terhadap suatu masalah, peristiwa, dan kegiatan masyarakat.

Analisis sentimen masyarakat terhadap Tapera di media sosial *X* akan dibahas dalam penelitian ini. Banyak pengguna media sosial yang membagikan pemikiran dan pendapatnya mengenai Kebijakan Tapera di platform media sosial *X*, berdasarkan peristiwa yang sedang terjadi di masyarakat. Hal ini menarik untuk diteliti guna mengetahui opini masyarakat tentang Tapera.

Sebuah algoritma diperlukan untuk mengkategorikan komentar publik di media sosial *X*, baik yang baik maupun yang buruk, untuk memudahkan penelitian tentang topik ini.

K-Nearest Neighbor adalah metode yang digunakan dalam penelitian ini untuk mengkategorikan komentar di media sosial *X*.

Karena beragamnya perspektif masyarakat terhadap Kebijakan Tapera, maka diperlukan sebuah perangkat lunak yang dapat mengkategorikan dukungan atau penolakan terhadap kebijakan tersebut. Oleh karena itu, peneliti menyarankan untuk menerapkan teknik *K-Nearest Neighbor* untuk model klasifikasi.

Penelitian pertama yang dilakukan oleh Arief Asro'i dan Herny Februariyanti (2022) dengan judul Analisis sentimen pengguna Twitter terhadap perpanjangan PPKM menggunakan metode *K-Nearest Neighbor*. Algoritma ini diterapkan pada 6408 data tweet yang dikumpulkan antara tanggal 1 Juli 2021 hingga 31 Desember 2021, dengan menggunakan kata kunci "PPKM", berdasarkan hasil penelitian yang telah dilakukan. Hasil training model menunjukkan bahwa nilai akurasi, recall, dan precision masing-masing sebesar 69.5%, 69.5%, dan 68.7%. Karena penulis membutuhkan presisi yang tinggi agar sah jika digunakan untuk evaluasi kebijakan, maka hasil yang didapatkan belum memenuhi standar. Dapat disimpulkan dari penelitian ini bahwa ada lebih banyak tweet dari pengguna Twitter yang netral atau tidak memiliki sentimen. Pengguna Twitter lebih cenderung memiliki sentimen negatif terhadap aturan PPKM jika sentimen netral diabaikan. karena persentase bulanan sentimen positif tidak pernah naik di atas 50%. (Arief Asro'i dan Herny Februariyanti, 2022).

Penelitian kedua dilakukan oleh Luthfika, dkk (2021) dengan judul Analisis sentimen masyarakat terhadap *COVID-19* Pada Media Sosial twitter. Dalam penelitian ini, model analisis sentimen terhadap tweet *COVID-19* dari pengguna twitter dibangun dengan menggunakan algoritma Naive Bayes dan K-Nearest Neighbor. Untuk $k = 6, 8, \text{ dan } 14$, algoritma K-Nearest Neighbor menghasilkan akurasi sebesar 82%. Perayapan data dan pelabelan adalah dua prosedur yang dilakukan selama tahap pengumpulan data. Pustaka Tweepy dan bahasa pemrograman Python digunakan untuk melakukan prosedur perayapan data. Tujuan dari prosedur ini adalah untuk mendapatkan data tweet dari pengguna media sosial terkait *COVID-19* di Twitter. 'COVID' dan tagar '#COVID19' digunakan untuk mendapatkan data. Selain itu, 1000 titik data dipilih dari jumlah total data tweet yang dikumpulkan untuk melakukan analisis positif dan negatif. (Luthfika, dkk. 2021).

Penelitian ketiga dilakukan oleh Supriyanto, dkk (2023) Penerapan algoritma *K-Nearest Neighbor* (K-NN) untuk analisis sentimen publik terhadap pembelajaran daring. 1825 data tweet berbahasa Indonesia digunakan dalam analisis sentimen pengguna Twitter terhadap topik-topik yang berkaitan dengan kebijakan pemerintah terkait pembelajaran daring. Dengan menggunakan modul

Tweepy Python, nilai sentimen diklasifikasikan ke dalam dua kelompok positif dan negatif dengan menerapkan pembobotan TF-IDF pada kata-kata. Setelah dilakukan pengujian dengan K hingga 20, hasil penelitian menunjukkan bahwa K=10 menghasilkan nilai akurasi tertinggi, dengan nilai akurasi 84.65%, presisi mencapai 87%, recall 86%, f measure 87%, dan error rate mencapai 0.12%. Selain itu, hasil penelitian juga menunjukkan bahwa sikap masyarakat terhadap pembelajaran online cenderung positif. (Supriyanto, dkk. 2023).

2. RUANG LINGKUP

Dalam penelitian ini permasalahan mencakup:

1. Cakupan permasalahan

Berdasarkan uraian dari latar belakang, cakupan permasalahan yang akan dikaji dalam penelitian ini adalah Bagaimana Analisis Sentimen Masyarakat Terhadap Tapera Melalui Media Sosial X Menggunakan Metode *K-Nearest Neighbor*.

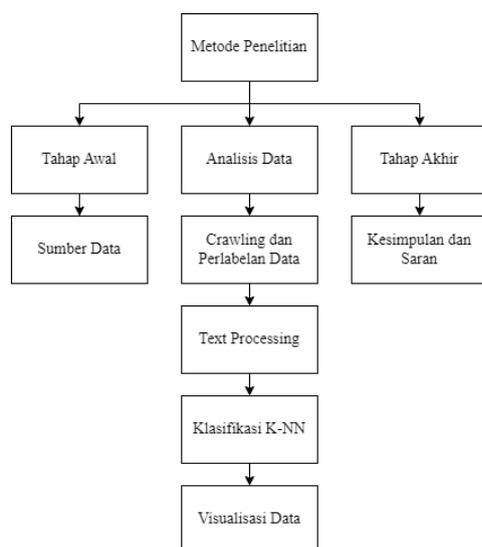
2. Batasan-batasan penelitian

Batasan masalah dalam penelitian ini adalah sebagai berikut :

- Data yang digunakan dalam penelitian ini adalah *Tweet* berbahasa Indonesia.
 - Menggunakan data sebanyak 958 *Tweet*
 - Metode yang digunakan untuk pengklafikasian dalam penelitian ini adalah metode *K-Nearest Neighbor*
 - Nilai K yang diujikan yaitu K=3, K=5 dan K=7
 - Sentimen yang akan diklasifikasikan yaitu sentimen positif, netral dan negatif
- ### 3. Rencana hasil yang didapatkan

Tujuan dari penelitian ini adalah untuk mengimplementasikan algoritma *K-Nearest Neighbor* (K-NN) dalam analisis sentimen pengguna media sosial X dengan topik tentang Tapera dan kemudian mengetahui nilai akurasi dan presisi dari metode tersebut.

3. BAHAN DAN METODE



Gambar 1 Alur Penelitian

3.1 Tahap Awal

1. Sumber Data

Data yang diperoleh merupakan sumber yang diambil dari X. Data yang digunakan berupa data *tweet* atau cuitan dari pengguna media sosial X di Indonesia berdasarkan kata kunci yang telah ditentukan.

3.2 Analisis Data

1. Crawling dan Pelabelan Data

Crawling dan pelabelan adalah dua tahapan yang akan dilakukan. Peneliti memperoleh data dari media sosial X menggunakan prosedur yang disebut *crawling* yang dimaksudkan untuk mengumpulkan informasi untuk dipelajari. Proses pelabelan kemudian menjadi prosedur kedua yang diselesaikan. Proses pelabelan melibatkan pemberian label positif, netral, dan negatif pada data yang telah dikumpulkan atau diperoleh melalui proses *crawling*.

2. Text Processing

Pada tahap ini ada tiga langkah yang akan diselesaikan. Pada tahap ini, data yang telah dikumpulkan, diakuisisi, dan diberi label akan diproses.

Case folding adalah prosedur awal yang dilakukan. Selama prosedur ini, semua huruf dokumen akan diubah menjadi huruf kecil, dan semua karakter selain huruf akan dihilangkan.

Selanjutnya, prosedur Tokenizing akan dijalankan, yang memerlukan pemotongan atau pemisahan setiap kata yang ada di dalam dokumen, yang menghasilkan token.

Filtering dilakukan setelah proses tokenizing. Selama prosedur ini, setiap kata yang tidak perlu seperti “dan”, “di”, “oleh”, dan seterusnya akan dihilangkan dari dokumen, dan hanya menyisakan istilah-istilah yang bermakna..

3. Klasifikasi *K-Nearest Neighbor*

Algoritma *K-Nearest Neighbor* akan digunakan untuk menangani data yang telah berhasil menyelesaikan tahap pelabelan, preprocessing, dan pembobotan metode TF-IDF. Selama tahap ini, komputer akan belajar mengidentifikasi pola dalam data dan mengklasifikasikannya ke dalam tiga kategori: kelas positif, netral, dan negatif.

4. Visualisasi Data

Pada tahapan visualisasi data akan menggunakan Presentase *Diagram Pie* dan *Word cloud* yang akan menampilkan data dari sentimen positif, netral dan negatif.

Tahap Akhir

1. Kesimpulan dan Saran

Berdasarkan dari proses analisis yang telah dilakukan dalam penelitian sehingga dapat memberikan kesimpulan dan saran penelitian.

4 HASIL DAN PEMBAHASAN

4.1 Sumber Data

Pada tahap ini yang diperoleh merupakan sumber yang diambil dari X. Data yang digunakan berupa data *tweet* atau cuitan dari pengguna media sosial X di Indonesia. Kata kunci untuk mencari data yang digunakan yaitu "tapera".

4.2 Crawling dan Pelabelan Data

Crawling data adalah proses pengumpulan informasi atau data dari berbagai sumber secara otomatis menggunakan program komputer atau bot. Proses ini mirip dengan cara mesin pencari seperti Google mengumpulkan informasi dari berbagai situs web untuk diindeks.

Pada tahap ini dalam proses pengambilan data menggunakan bahasa pemrograman *Python* dan menggunakan platform penulisan pemrograman *Google Colab* dan *tools* yang di gunakan untuk *Crawling* data adalah *tweet-harvest*. *Tweet-harvest* merupakan tools yang digunakan untuk melakukan *crawling* data pada media sosial X dengan menggunakan Application Programming Interface (API).

Pelabelan data adalah proses menandai atau menentukan sentimen dari teks atau data tertentu sebagai positif, negatif, atau netral. Proses pemberian label pada sebagian data secara manual sebanyak 350 data, lalu sebagian data yang lain dilakukan pelabelan secara otomatis.

Tabel 1. Data Tweet dan Labeling

id	<i>Tweet</i>	Polarity
1	@halahvator Gue nyesel semalem nonton berita ttg Tapera itu berasa kek beban rakyat ada di tangan gue sampe pusing dan ga mood. Padahal gue lagi nganggur malah mikirin sistem pemerintahan yg makin ngawur	negatif
2	Masalahnya gw udh punya rumah sendiri bambang. Kenapa maksa buat nabung ? Mending duitnya gw puterin biar ga kegerus inflasi. Ini kok semua lini pemerintah mencoba menjelaskan TAPERA ? Ada apa ?	negatif
3	TAPERA bukanlah Produk Undang ² Pro Rakyat. Dari persepsi manapun di kaji tidak akan mungkin terwujud untuk memiliki rumah dari hasil tabungan tersebut.	negatif
4	@CNNIndonesia Tapera sifatnya tabungan? Saya bisa mengelola keuangan sendiri pak jadi negara ga usah repot2	netral
5	@ARSIPAJA Punya kok solusinya yaitu dengan adanya tapera maka beberapa lowongan kerja baru terbuka	positif
...
958	@5teV3n_Pe9eL Dikira semua orang pengen punya rumah subsidi. Pdahal bangun sendiri pelan pelan dari uang gaji jauh lebih baik daripada Tapera	negatif

4.3 Text Processing

Text Processing dilakukan selanjutnya, setelah pelabelan data sebagai positif, netral, atau negatif. Saat ini, sejumlah prosedur akan diterapkan pada data yang tersedia, termasuk case folding, tokenizing dan filtering. Case Folding merupakan prosedur pertama yang dilakukan pada tahap ini. Semua huruf kapital (huruf besar) dalam data akan diubah menjadi huruf kecil selama prosedur case folding.

Tabel 2. Proses case folding

Sebelum Case folding	Sesudah Case folding
Gak sekalian nih tiap orderan makan daring belanja daring tagihan internet listrik pdam bpjs kesehatan sppukt sewa lapak rumah apartemen iuran RTRW dll langsung dipotong Tapera	gak sekalian nih tiap orderan makan daring belanja daring tagihan internet listrik pdam bpjs kesehatan sppukt sewa lapak rumah apartemen iuran rtrw dll langsung dipotong tapera
INI INDONESIA KEKURANGAN TAPERA SALAH SATU LAHAN EMPUK UNTUK DIMAINKAN ATAU TIDAKNYA KITA LIHAT SAJA NANTINYA	ini indikasi indonesia mulai kekurangan duit tapera adalah salah satu lahan empuk untuk dimainkan betul atau tidaknya kita lihat saja nantinya

Tokenizing menjadi langkah berikutnya setelah Case folding. Setiap kata dalam teks dibagi menjadi token selama prosedur ini.

Tabel 3. Proses Tokenizing

Sebelum Tokenize	Sesudah Tokenize
jika program tapera memaksa mengutip uang dari pekerja swasta pengemudi ojol niscaya akan terjadi gejolak sosial	'jika', 'program', 'tapera', 'memaksa', 'mengutip', 'uang', 'dari', 'pekerja', 'swasta', 'pengemudi', 'ojol', 'niscaya', 'akan', 'terjadi', 'gejolak', 'sosial'
saat ini rumah kpr subsidi harga juta dengan cicilan jutaan perbulan dalam jangka waktu tahun kalau tapera ribu perbulan kapan bisa kebeli rumahnya	'saat', 'ini', 'rumah', 'kpr', 'subsidi', 'harga', 'juta', 'dengan', 'cicilan', 'jutaan', 'perbulan', 'dalam', 'jangka', 'waktu', 'tahun', 'kalau', 'tapera', 'ribu', 'perbulan', 'kapan', 'bisa', 'kebeli', 'rumahnya'

Setelah dilakukan *Tokenizing*, prose terakhir adalah filtering stopword. Pada proses filtering stopword dilakukan penghapusan kosakata tidak bermakna yang terdapat dalam dokumen. Kosakata tak bermakna meliputi "dan", "di", "oleh", "karena", "yang", "ini", dan lain sebagainya.



Tabel 4. Proses Filtering

Sebelum Filtering	Sesudah Filtering
indonesia sedang tidak baik saja hutang banyak pembangunan ikn butuh modal banyak investor tidak ada akhirnya puter otak dengan cara memeras rakyat dengan program tapera tabungan peras rakyat nya bravo bravo hebat kalian semua pejabat nkri	indonesia hutang pembangunan ikn butuh modal investor akhirnya puter otak memeras rakyat program tapera tabungan peras rakyat bravo bravo hebat pejabat nkri

4.4 Klasifikasi K-Nearest Neighbor

1. Pembobotan Kata (TF-IDF)

Dokumen yang siap untuk diproses pada langkah berikut, yaitu pembobotan kata setelah menyelesaikan semua langkah yang terlibat dalam prapemrosesan teks. Pada tahap ini, frekuensi atau jumlah kemunculan kata dalam dokumen akan dihitung dengan perhitungan. Secara keseluruhan, dari total 958 *tweet*, terdapat 8781 term yang berbeda.

Tabel 5. Pembobotan kata dengan TF-IDF

Dokumen	Dokumen								
	program tapera memaksa mengutip uang pekerja swasta pengemudi ojol niscaya gejala sosial								
	tapera pungutan diluar pajak berkedok tabungan								
Term	tf			df	d/df	IDF	w		
	d1	d2	d3				d1	d2	d3
	berkedok	0	1	0	1	3	0,48	0	0,48
gejala	1	0	0	1	3	0,48	0,48	0	0
memaksa	1	0	0	1	3	0,48	0,48	0	0
mengutip	1	0	0	1	3	0,48	0,48	0	0
niscaya	1	0	0	1	3	0,48	0,48	0	0
ojol	1	0	0	1	3	0,48	0,48	0	0
pajak	0	1	0	1	3	0,48	0	0,48	0
pekerja	1	0	0	1	3	0,48	0,48	0	0
pemerasan	0	0	1	1	3	0,48	0	0	0,48
pengemudi	1	0	0	1	3	0,48	0,48	0	0
program	1	0	0	1	3	0,48	0,48	0	0
pungutan	0	1	0	1	3	0,48	0	0,48	0
rakyat	0	0	1	1	3	0,48	0	0	0,48
sosial	1	0	0	1	3	0,48	0,48	0	0
swasta	1	0	0	1	3	0,48	0,48	0	0
tabungan	1	1	0	1	1,5	0,18	0,18	0,18	0
uang	1	0	0	1	3	0,48	0,48	0	0

2. Pengujian dan Evaluasi

Pada tahap pengujian dan evaluasi dalam penelitian yang dilakukan ini adalah menggunakan model data training yang sudah diberi label secara manual sebanyak 349 data. Model training akan diuji dengan model uji sebanyak 609 data *tweet*.

Tabel 6. Sampel hasil prediksi

id	tweet	label aktual	label prediksi
1	menyala pemerintahku dengan tapera yang dibarengi dengan penghapusan minimal umur calon kepala daerah nya	negatif	netral
2	@ernestprakasa tapera ini dibentuk sejatinya untuk dipersiapkan dan dipakai nantinya buat bancakan korupsi pejabat? di masa datang.	negatif	negatif
3	@ARSIPAJA Punya kok solusinya yaitu dengan adanya tapera maka beberapa lowongan kerja baru terbuka	positif	netral

Sampel dari hasil prediksi, kemudian akan direpresentasikan ke dalam *confusion matrix*. Representasi *confusion matrix* untuk K=3 dapat dilihat pada tabel berikut

Tabel 7. Confusion matrix pengujian K=3

		Nilai aktual		
		netral	positif	negatif
Nilai Prediksi	netral	31	4	126
	positif	0	1	0
	negatif	0	0	187

Berdasarkan Tabel 7. *Confusion matrix* pengujian K=3, maka perolehan nilai akurasi, presisi dan recall dapat dilihat pada tabel berikut.

Tabel 8. Confusion matrix Formula

Confusion matrix Formula	
Akurasi	$\frac{\text{Number of Correct Prediction}}{\text{Total Number of Prediction}}$
Presisi	netral = $\frac{TNe}{TNe+FNe}$
	positif = $\frac{TP}{TP+FP}$
	negatif = $\frac{TN}{TN+FN}$
Recall	netral = $\frac{TNe}{TNe+FP+FN}$
	positif = $\frac{TP}{TP+FNe+FN}$
	negatif = $\frac{TN}{TN+FP+FNe}$

Tabel 9. Nilai pengujian K=3

Pengujian		
Akurasi	$\frac{31+1+187}{31+4+127+0+1+0+0+0+187} = \frac{219}{349}$	62,75%
Presisi	netral = $\frac{31}{31+130}$	19,25%
	positif = $\frac{1}{1}$	100%
	negatif = $\frac{187}{187}$	100%
Recall	netral = $\frac{31}{31}$	100%
	positif = $\frac{1}{5}$	20%
	negatif = $\frac{187}{187+126}$	59,74%

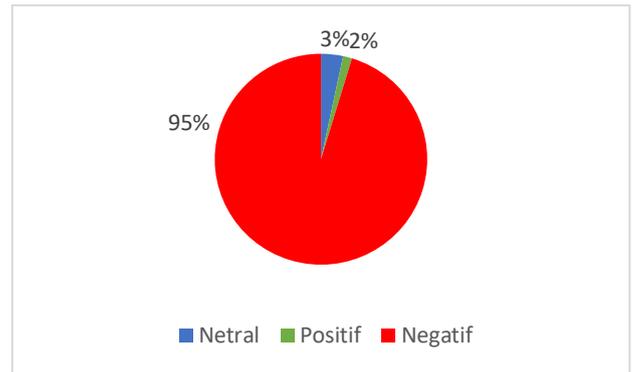
Pengujian di atas dilakukan secara berulang dengan variasi nilai K yang diujikan yaitu, K=3, K=5 dan K=7. Sehingga dapat diketahui hasil pengujian secara keseluruhan.

Tabel 10. Hasil pengujian dan evaluasi

		K=3	K=5	K=7
Akurasi		0,6275	0,4069	0,3438
Presisi	Netral	0,1925	0,1140	0,1163
	Negatif	1	0	0
Recall	Netral	1	0,9587	0,9890
	Positif	0,20	0	0
	Negatif	0,5974	0,37	0,2875

4.5 Visualisasi Data

Pada tahap ini adalah melakukan visualisasi data dengan menggunakan *Diagram Pie* dan *word cloud*. Visualisasi data dengan *Diagram Pie* menggambarkan perbandingan presentasi sentimen positif, sentimen netral dan sentimen negatif. Sedangkan untuk penggunaan *word cloud* ialah menampilkan banyaknya sebuah kata yang tampil pada setiap dokumen.



Gambar 2. Visualisasi Diagram Pie

Berdasarkan penelitian yang telah dilakukan, dapat diketahui bahwa sebanyak 95% *tweet* negatif, 3% *tweet* netral dan 2% *tweet* positif dengan total data sebanyak 958 *tweet*.



Gambar 3. Visualisasi word cloud

5. KESIMPULAN

Dari hasil pengujian perbandingan antara *tweet* yang di tuliskan oleh masyarakat terhadap Tapera yaitu cenderung ke arah sentimen yang buruk (negatif) sebesar 95%. Dataset yang digunakan sebanyak 958 *tweet*, dimana 911 data memiliki sentimen negatif, 34 data memiliki sentimen netral dan 13 data memiliki sentimen positif. Dari Hasil pengujian menggunakan *confusion matrix* dari Algoritma K Nearest Neighbor, diketahui bahwa nilai akurasi terendah diperoleh dengan nilai K=7 dengan akurasi 34,38%, sedangkan nilai akurasi tertinggi didapatkan dengan nilai K=3 dengan akurasi 62,75%

6. SARAN

Disarankan penelitian ini dapat dimulai dengan mengumpulkan data *tweet* terkait Tapera melalui scraping data twitter. Lakukan pemrosesan data dengan tokenisasi. Implementasikan metode K-NN untuk klasifikasi sentimen.



7. DAFTAR PUSTAKA

- Afriyanti, A. (2022). Analisis Sentimen Masyarakat Mengenai Vaksin Covid-19 Di Indonesia Menggunakan Algoritma *K-Nearest Neighbor* (KNN) dan Naïve Bayes. <https://repository.unilak.ac.id/2830/>
- Asro'i, A., & Februariyanti, H. (2022). Analisis Sentimen Pengguna Twitter Terhadap Perpanjangan Ppkm Menggunakan Metode *K-Nearest Neighbor*. *Jurnal Khatulistiwa Informatika*, 10(1), 17-24. <https://doi.org/10.31294/jki.v10i1.12624>
- Astuti, P., & Nuris, N. (2022). Penerapan Algoritma KNN Pada Analisis Sentimen Review Aplikasi Peduli Lindungi. *Computer Science (Co-Science)*, 2(2), 137-142. <https://doi.org/10.31294/coscience.v2i2.1258>
- Baita, A., Pristyanto, Y., & Cahyono, N. (2021). Analisis sentimen mengenai vaksin sinovac menggunakan algoritma support vector machine (SVM) dan *K-Nearest Neighbor* (KNN). *Information System Journal*, 4(2), 42-46. <https://doi.org/10.24076/infosjournal.2021v4i2.687>
- Fairuz, A. L., Ramadhani, R. D., & Tanjung, N. A. F. (2021). Analisis Sentimen Masyarakat Terhadap COVID-19 Pada Media Sosial Twitter. *Journal of Dinda: Data Science, Information Technology, and Data Analytics*, 1(1), 42-51. <https://doi.org/10.20895/dinda.v1i1.180>
- Febrian, Z. (2023). Perbandingan Algoritma Klasifikasi Naïve Bayes, Logistic Regression, Dan KNN Untuk Analisis Sentimen Peminatan Masyarakat Terhadap Kandidat Bakal Calon Presiden 2024. <https://repository.ar-raniry.ac.id/id/eprint/34820>
- Latif, a. (2021). Perbandingan kinerja akurasi pada analisis sentimen terhadap vaksinasi covid-19 pada twitter menggunakan algoritma nave bayes dan support vector machine. <https://repository.mercubuana.ac.id/id/eprint/84566>
- Pamungkas, F. S., & Kharisudin, I. (2021, February). Analisis Sentimen dengan SVM, NAIVE BAYES dan KNN untuk Studi Tanggapan Masyarakat Indonesia Terhadap Pandemi Covid-19 pada Media Sosial Twitter. In *PRISMA, Prosiding Seminar Nasional Matematika* (Vol. 4, pp. 628-634). <https://journal.unnes.ac.id/sju/prisma/article/view/45038>
- Prasetyo, S. D., Hilabi, S. S., & Nurapriani, F. (2023). Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma Naïve Bayes dan KNN. *Jurnal KomtekInfo*, 1-7. <https://doi.org/10.35134/komtekinfo.v10i1.330>
- Pratama, A. Y., & Voutama, A. (2021). Analisis Sentimen Media Sosial Twitter Dengan Algoritma *K-Nearest Neighbor* Dan Seleksi Fitur Chi-Square (Kasus Omnibus Law Cipta Kerja). *J-SAKTI (Jurnal Sains Komputer dan Informatika)*, 5(2), 897-910. <http://www.tunasbangsa.ac.id/ejurnal/index.php/jskti/article/view/386>
- Pratomo, S. A., Al Faraby, S., & Purbolaksono, M. D. (2021). Analisis Sentimen Pengaruh Kombinasi Ekstraksi Fitur TF-IDF dan Lexicon Pada Ulasan Film Menggunakan Metode KNN. <https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/15726?btwaf=36235012>
- Putra, A. D. A., & Juanita, S. (2021). Analisis Sentimen pada Ulasan pengguna Aplikasi Bibit Dan Bareksa dengan Algoritma KNN. *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, 8(2), 636-646. <https://doi.org/10.35957/jatisi.v8i2.962>
- Putra, R. A., & Pramusinto, W. (2023, October). PENERAPAN ALGORITMA *K-NEAREST NEIGHBOR* (KNN) ANALISIS SENTIMEN TWEET PADA ELEKTABILITAS BAKAL CALON PRESIDEN 2024. In *Prosiding Seminar Nasional Mahasiswa Fakultas Teknologi Informasi (SENAFTI)* (Vol. 2, No. 2, pp. 985-994). <https://senafti.budiluhur.ac.id/index.php/senafti/article/view/967>
- Rohmansyah, F. A., Bintoro, B., & Santoso, I. (2023). Analisis Sentimen Terhadap Penerapan Sistem Ganjil Genap Menggunakan Metode *K-Nearest Neighbor* (KNN). *IKRA-ITH Informatika: Jurnal Komputer dan Informatika*, 7(2), 165-169. <https://journals.upi-yai.ac.id/index.php/ikraith-informatika/article/view/2826>
- Supriyanto, J., Alita, D., & Isnain, A. R. (2023). Penerapan Algoritma *K-Nearest Neighbor* (K-NN) Untuk Analisis Sentimen Publik Terhadap Pembelajaran Daring. *Jurnal Informatika dan Rekayasa Perangkat Lunak*, 4(1), 74-80. <https://doi.org/10.33365/jatika.v4i1.2468>