



ANALISIS SENTIMEN MENGGUNAKAN K-NEAREST NEIGHBOR TERHADAP FILM NGERI-NGERI SEDAP

Shany Maranatha Sipayung¹, Margaretha Yohanna², Harlen Gilbert Manullang³, Rimky Mandala Putra Simanjuntak⁵

^{1,2,3}Fakultas Ilmu Komputer, Universitas Methodist Indonesia

⁴Fakultas Ekonomi, Universitas Methodist Indonesia

Info Artikel

Histori Artikel:

Received, Agus 9, 2024

Revised, Sep 20, 2024

Accepted, Sep 11, 2024

Keywords:

Twitter,
Ngeri-Ngeri Sedap,
Analisis Sentimen,
KNN,

ABSTRAK

Film Ngeri-Ngeri Sedap sempat viral dan menjadi perbincangan masyarakat Indonesia pada awal penayangannya. Setidaknya ada beberapa persepsi yang muncul setelah menyaksikan film tersebut, misalnya adat pada suku batak yang masih kental, adanya rasis skala kecil, orangtua yang merasa berhak atas anak, pesta orang batak selalu menghabiskan banyak biaya, dll. Dengan adanya persepsi tersebut sehingga menimbulkan opini masyarakat yang berbeda khususnya para pengguna twitter. Opini yang diunggah tersebut selanjutnya akan dijadikan data dari penelitian Penerapan Metode K-Nearest Neighbor Dalam Menganalisis Sentimen Masyarakat Pada Sosial Media Twitter Terhadap Film (Studi Kasus : Ngeri-Ngeri Sedap). Untuk pembobotan data dilakukan dengan menggunakan perhitungan TF-IDF. Dan untuk perhitungan jarak digunakan rumus Euclidean Distance. Untuk klasifikasi sentimen menggunakan metode K-Nearest Neighbor dan digolongkan kedalam kelas positif, kelas netral, dan kelas negatif. Tools yang digunakan pada penelitian ini adalah Rapid Miner. Pengukuran evaluation measure dengan rasio perbandingan 60% data training dan 40% data uji didapatkan hasil Precision sebesar 79%, Recall 100%, dan F1-score 88%.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Penulis Koresponden:

Shany Maranatha Sipayung
Fakultas Ilmu Komputer,
Universitas Methodist Indonesia, Medan,
Jl. Hang Tuah No.8, Medan - Sumatera Utara.
Email: shanimaranatha29@gmail.com

1. PENDAHULUAN

Film Ngeri-Ngeri Sedap adalah film yang mulai ditayangkan pada tanggal 2 Juni 2022 di seluruh bioskop Indonesia. Film Ngeri-Ngeri Sedap ini sangat viral dan menjadi perbincangan masyarakat Indonesia pada awal awal penayangannya. Penyajian filmnya sangat berkesinambungan dengan kehidupan keluarga yang ada di Sumatera Utara terutama di tanah Batak. Setidaknya ada beberapa persepsi yang muncul setelah menyaksikan film Ngeri-Ngeri Sedap ini, misalnya adat pada suku Batak yang masih kental, adanya rasis skala kecil, orangtua yang merasa berhak atas anak, pesta orang Batak selalu menghabiskan banyak biaya, anak perempuan yang selalu menjadi nomor kesekian, dan anak lelaki yang paling kecil memiliki tanggung jawab untuk menjaga orangtua di masa tuanya. Dengan adanya persepsi masyarakat terhadap film Ngeri-Ngeri Sedap ini, terbentuk sentimen masyarakat menjadi 3 sentimen, diantaranya ada masyarakat yang positif, negatif, dan netral terhadap penayangan film Ngeri-Ngeri Sedap ini. Hal ini terlihat dari komentar-komentar pada sosial media twitter dengan topik film Ngeri-Ngeri Sedap. Maka dari

itu, perlu adanya klasifikasi yang dilakukan terkait pandangan masyarakat terhadap film Ngeri-Ngeri Sedap ini, dan metode yang bisa digunakan dalam proses klasifikasi ini adalah *K-nearest neighbor* (KNN). Metode KNN adalah salah satu metode *supervised learning* yang proses klasifikasinya dilakukan berdasarkan jarak ketetanggaan terdekat dari dataset training [1]. Algoritma KNN ini adalah sebuah metode yang bisa digunakan untuk mengklasifikasikan objek berdasarkan data pelatihan yang paling dekat dengannya. Data pelatihan dideskripsikan dalam ruang multidimensi, dengan setiap dimensi mewakili setiap fitur data. Klasifikasi data baru dilakukan dengan mencari k label tetangga terdekat. Sebagian besar label yang Anda lihat adalah label data baru. Jika $k = 1$ maka data baru diberi label tetangga terdekat. Analisis sentimen ini dapat menentukan pendapat atau opini dari suatu teks atau dokumen kedalam kelas positif, netral, maupun negatif [2].

2. METODE PENELITIAN

2.1 Data Mining

Data mining atau penambangan data melibatkan penggalian pola, informasi, dan pengetahuan tersembunyi yang berharga dari kumpulan data yang besar dan kompleks.. Tujuannya yakni guna mengidentifikasi hubungan yang berguna, tren, pola, ataupun pengetahuan yang mungkin tidak nampak secara langsung lewat pengamatan biasa. Penambangan data adalah analisis yang meninjau kumpulan data, menciptakan hubungan yang tidak dapat diprediksi, dan merangkum data menggunakan metode berbeda yang dapat dipahami dan berguna bagi pemilik data. Data mining merupakan salah satu cabang dari beberapa bidang keilmuan yang menangani kasus pengambilan data dari database berukuran besar dengan menggabungkan metode pembelajaran mesin, pengenalan pola, statistik, database, dan visualisasi. Penambangan data adalah proses mengekstraksi dan mengidentifikasi informasi berguna dan pengetahuan terikat dari database besar menggunakan teknik statistik, matematika, kecerdasan buatan, dan pembelajaran mesin [3].

2.2 Klasifikasi

Klasifikasi merupakan sebuah metode pengelompokan data yang telah mempunyai kelas sebelumnya. Dengan menggunakan data-data histori, akan terbentuk sebuah rule yang dapat digunakan untuk menentukan kelas dari data berikutnya [4]. Tujuan utama klasifikasi adalah untuk mengidentifikasi pola atau hubungan yang ada dalam data sehingga data-data yang serupa dapat dikelompokkan ke dalam kelas yang sama. Klasifikasi adalah salah satu tugas penting dalam analisis data dan pembelajaran mesin. Dalam konteks data mining, klasifikasi melibatkan penggunaan model statistik atau algoritma pembelajaran mesin untuk membuat prediksi atau keputusan kelas berdasarkan data pelatihan yang telah diberikan.

2.3 Text Mining

Text Mining adalah proses pengumpulan informasi secara intensif menggunakan alat dan metode khusus yang digunakan untuk menganalisis data dan dokumen. Text Mining adalah bagian dari penambangan data dan dapat menganalisis data semi terstruktur (Word, PDF, kutipan teks) dan data tidak terstruktur.

Tujuan dari penambangan teks adalah untuk memahami dan mengambil informasi berguna dari sumber data dengan mengidentifikasi dan mengeksplorasi pola bahasa yang unik. Dalam kasus text mining, sumber data yang digunakan adalah collection atau unstructured collection dan memerlukan kategorisasi untuk menemukan informasi sejenis [1]. Text mining adalah penerapan konsep dan teknik data mining. Namun, karena penambangan teks melibatkan pemrosesan data teks tidak terstruktur, penambangan teks memiliki lebih banyak tahapan dibandingkan penambangan data. Berdasarkan hal ini, kita memerlukan langkah pertama untuk menyiapkan data teks untuk diproses: prapemrosesan.

2.4 Preprocessig Text

Preprocessing teks adalah serangkaian langkah atau teknik yang digunakan untuk membersihkan, menyiapkan, dan mengubah teks mentah menjadi format yang lebih sesuai untuk analisis teks atau pemrosesan bahasa alami (NLP). Pada umumnya, preprocessing data dilakukan dengan cara mengeliminasi data yang tidak sesuai atau mengubah data menjadi bentuk yang lebih mudah diproses oleh sistem [5]. Tujuan dari preprocessing teks adalah untuk meningkatkan kualitas data teks, menghilangkan noise, dan memastikan bahwa teks siap digunakan dalam berbagai aplikasi analisis teks, termasuk analisis sentimen, klasifikasi teks, ekstraksi informasi, dan lain-lain. *Preprocessing text* terdiri dari beberapa tahapan yaitu :

1. Cleansing, berfungsi untuk membersihkan atribut yang tidak berpengaruh seperti simbol, angka, link, dll.
2. Case Folding, berfungsi mengubah seluruh huruf dalam suatu dokumen menjadi huruf kecil
3. Tokenizing, adalah tahap memotong string input berdasarkan setiap kata yang menyusunnya
4. Filtering, adalah tahap mengambil kata-kata penting dari hasil tokenizing
5. Stemming, selain untuk mengurangi jumlah indeks yang berbeda-beda dalam suatu dokumen, juga digunakan untuk mengelompokkan kata-kata lain yang mempunyai kata dasar dan makna yang sama, namun bentuknya berbeda karena mempunyai imbuhan yang berbeda

2.5 Term Frequency-Inverse Document Frequency (TF-IDF)

Term Frequency-Inverse Document Frequency (TF-IDF) adalah metode algoritmik yang memberi bobot pada setiap kata dan biasa digunakan sebagai strategi untuk mengklasifikasikan teks atau dokumen. TF-IDF adalah ukuran statistik yang menggambarkan pentingnya istilah dalam dokumen dalam suatu koleksi atau korpus. Ukuran ini sering digunakan sebagai faktor pembobotan dalam pengambilan informasi, penambangan teks, dan pemodelan pengguna. Nilai tf-idf meningkat secara linier dengan banyaknya kemunculan suatu term dan bergantung pada jumlah dokumen dalam korpus yang memuat istilah tersebut. Istilah frekuensi (tf) menunjukkan frekuensi (derajat frekuensi) suatu istilah dalam suatu dokumen. Sedangkan frekuensi dokumen (df) adalah banyaknya dokumen yang memunculkan istilah. TF-IDF adalah perhitungan atau pembobotan kata melalui teknik tokenisasi, stopwords, dan stemming, dan frekuensi munculnya kata dalam dokumen yang diberikan menunjukkan pentingnya kata itu di dalam sebuah dokumen [6].

Perhitungan TF-IDF dapat dilakukan dengan menggunakan rumus sebagai berikut.

$$\text{TF-IDF} = \text{Term frequency pada setiap dokumen} \times \text{Inverse Document frequency}$$

2.6 K-Nearest Neighbor

K-nearest neighbor (KNN) adalah metode yang mudah untuk digunakan dalam melakukan klasifikasi teks. KNN adalah salah satu metode *supervised learning* yang proses klasifikasinya berdasarkan jarak ketetanggaan terdekat dari dataset training [1]. K-NN merupakan salah satu metode non parametrik yang prinsip kerjanya adalah mengklasifikasikan objek pada set pengujian berdasarkan kelas terbaik dari k tetangga terdekat pada data pelatihan. Nilai k merupakan bilangan bulat positif yang ditentukan secara acak oleh pengguna. Untuk mendapatkan nilai k yang baik dapat dipilih dengan optimasi parameter [7] . Pada fase training, algoritma ini hanya melakukan penyimpanan vektor-vektor fitur dan klasifikasi data training sample [8]. Klasifikasi teks menggunakan metode KNN akan menghasilkan nilai yang lebih optimal jika menggunakan rumus *Euclidean Distance* untuk pembobotan tiap-tiap kata pada dokumen teks

yang akan diproses. Untuk menghitung nilai jarak antar dokumen menggunakan Euclidean Distance dapat menggunakan rumus sebagai berikut.

$$\text{Euclidean Distance}(A, B) = \sum_{i=1}^t \sqrt{(A - B)^2}$$

Dimana :

- A : dokumen testing atau uji
 B : dokumen training atau latihan
 t : jumlah term atau kata

2.7 Evaluation Measure

Evaluasi dalam sebuah algoritma bertujuan untuk menilai performansi yang dapat dicapai oleh sistem. Akurasi adalah kriteria paling umum untuk mengukur kinerja klasifikasi. Namun, kriteria ini tidak terlalu tepat ketika berhadapan dengan kelas-kelas yang tidak seimbang, karena kelas-kelas minoritas hanya memberikan kontribusi yang kecil terhadap kriteria akurasi. Kriteria Penilaian yang disarankan adalah F1-score untuk mengukur klasifikasi kelas minoritas pada kelas tidak seimbang [9]. Dalam mengukur performansi suatu metode, F1-score digunakan untuk melakukan suatu perhitungan evaluasi yang mengkombinasikan nilai antara precision dan recall [10].

Precision merupakan parameter yang digunakan untuk mengukur keakuratan suatu algoritma. Misalnya, untuk menghitung keakuratan algoritme yang memprediksi data berlabel positif, Anda dapat menghitung keakuratan berdasarkan rasio jumlah data berlabel positif yang diprediksi dengan benar oleh algoritme terhadap jumlah data yang diperkirakan berlabel positif oleh algoritme. Untuk menghitung precision dapat menggunakan rumus sebagai berikut.

$$\text{Precision} = \frac{TP}{TP+FP}$$

Recall merupakan salah satu cara untuk menghitung kelengkapan suatu algoritma. Misalnya, untuk menghitung recall suatu algoritma dalam memprediksi data berlabel negatif, Anda perlu menghitung jumlah data berlabel negatif yang diprediksi dengan benar oleh algoritma dan jumlah data berlabel negatif dalam dataset dihitung berdasarkan rasio jumlah data terlampir. Untuk menghitung recall dapat menggunakan rumus sebagai berikut.

$$\text{Recall} = \frac{TP}{TP + FN}$$

F1-score merupakan rata-rata harmonik dari presisi dan recall. Nilai tertinggi adalah 1 dan nilai terendah adalah 0. Untuk menghitung F1-score dapat menggunakan rumus sebagai berikut.

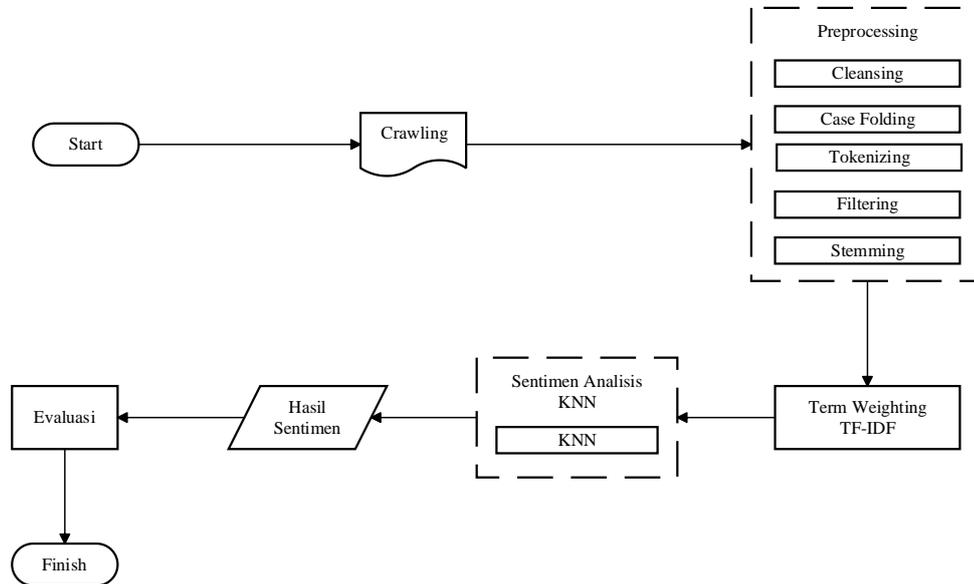
$$F1 - \text{Score} = 2 * \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

2.8 Flowchart

Flowchart adalah rangkaian tahapan penyelesaian masalah yang di tuliskan dalam bentuk simbol-simbol tertentu. Flowchart adalah gambaran diagram atau bentuk urutan proses dalam sebuah program yang terhubung antara proses dengan tampilannya, sehingga mengalir pada suatu prosedur yang sudah dirancang [11].

2.9 Framework Penelitian

Penelitian mengenai klasifikasi analisa sentimen ini memiliki rancangan penelitian yang akan dikerjakan. Gambaran umum dari penelitian ini dapat dilihat seperti gambar 1 sebagai berikut.



Gambar 1. Framework Penelitian

3. HASIL DAN PEMBAHASAN

3.1 Implementasi

Tahapan pengujian sistem menggunakan data komentar dari twitter, menggunakan 184 data latih dan 46 data uji. Untuk kategori komentar yang digunakan ada sebanyak 3 kelas yaitu komentar positif, netral, dan negatif.

3.1.1 Preprocessing Text

Preprocessing bertujuan untuk mempersiapkan teks yang belum terstruktur menjadi teks yang terstruktur sehingga dapat digunakan untuk proses selanjutnya.

Proses cleansing berfungsi untuk menghapus simbol, tanda baca, angka, url, dll. Proses cleansing pada data tweet dapat dilihat pada tabel 1 sebagai berikut

Tabel 1. Tabel Cleansing

Doc	Sebelum Cleansing	Sesudah Cleansing
Doc1	ngeri ngeri sedap adalah film yg sangat layak ditonton	ngeri ngeri sedap adalah film yg sangat layak ditonton
Doc2	film indo yang bener bener bagus buat gua itu ngeri ngeri sedap~	film indo yang bener bener bagus buat gua itu ngeri ngeri sedap
Doc3	Liat masku pamit mau kerja ke tempat jauh	Liat masku pamit mau kerja ke tempat jauh
Doc4	OIYAAA FILM NGERI NGERI SEDAP JUGA!!!! asli itu nangis jelek banget pas dr tengah ke akhir ðŸ˜	OIYAAA FILM NGERI NGERI SEDAP JUGA asli itu nangis jelek banget pas dr tengah ke akhir

Doc	Sebelum Cleansing	Sesudah Cleansing
Doc5	Film ngeri ngeri sedap bikin gua nangis	Film ngeri ngeri sedap bikin gua nangis

Proses case folding berfungsi untuk merubah huruf kapital menjadi huruf kecil. Proses case folding pada data tweet dapat dilihat pada tabel 2 sebagai berikut.

Tabel 2. Tabel Case Folding

Doc	Sebelum Case Folding	Sesudah Case Folding
Doc1	ngeri ngeri sedap adalah film yg sangat layak ditonton	ngeri ngeri sedap adalah film yg sangat layak ditonton
Doc2	film indo yang bener bener bagus buat gua itu ngeri ngeri sedap	film indo yang bener bener bagus buat gua itu ngeri ngeri sedap
Doc3	Liat masku pamit mau kerja ke tempat jauh	liat masku pamit mau kerja ke tempat jauh
Doc4	OIYAAA FILM NGERI NGERI SEDAP JUGA asli itu nangis jelek banget pas dr tengah ke akhir	oiyaaa film ngeri ngeri sedap juga asli itu nangis jelek banget pas dr tengah ke akhir
Doc5	Film ngeri ngeri sedap bikin gua nangis	film ngeri ngeri sedap bikin gua nangis

Proses tokenizing berfungsi untuk memisahkan kalimat menjadi kata per kata. Proses tokenizing pada data tweet dapat dilihat pada tabel 3 sebagai berikut.

Tabel 3. Tabel Tokenizing

Doc1	Doc2	Doc3	Doc4	Doc5
ngeri	film	liat	oiyaaa	film
ngeri	indo	masku	film	ngeri
sedap	yang	pamit	ngeri	ngeri
adalah	bener	mau	ngeri	sedap
film	bener	kerja	sedap	bikin
yg	bagus	ke	juga	gua
sangat	buat	tempat	asli	nangis
layak	gua	jauh	itu	
ditonton	itu		nangis	
	ngeri		jelek	
	ngeri		banget	
	sedap		pas	
			dr	
			tengah	
			ke	
			akhir	

Proses filtering berfungsi untuk menghilangkan kata kata yang tidak berpengaruh. Proses filtering pada data tweet dapat dilihat pada tabel 4 sebagai berikut.

Tabel 4. Tabel Filtering

Doc1	Doc2	Doc3	Doc4	Doc5
ngeri	film	liat		film
ngeri	indo	masku	film	ngeri
sedap	yang	pamit	ngeri	ngeri
adalah	bener	mau	ngeri	sedap

Doc1	Doc2	Doc3	Doc4	Doc5
film	bener bagus	kerja	sedap	bikin gua
sangat layak ditonton	buat gua ngeri ngeri sedap	tempat jauh	asli nangis jelek banget pas tengah akhir	nangis

Proses pada stemming berfungsi untuk menjadikan suatu kata ke bentuk kata dasar. Proses stemming pada data tweet dapat dilihat pada tabel 5 sebagai berikut.

Tabel 5. Tabel Stemming

Doc1	Doc2	Doc3	Doc4	Doc5
ngeri	film	liat	film	film
ngeri	indo	masku	ngeri	ngeri
sedap	yang	pamit	ngeri	ngeri
adalah	bener	mau	sedap	sedap
film	bener	kerja	asli	bikin
sangat	bagus	tempat	nangis	gua
layak	buat	jauh	jelek	nangis
tonton	gua		banget	
	ngeri		pas	
	ngeri		tengah	
	sedap		akhir	

3.1.2 Pembobotan Dokumen (TF-IDF)

Setelah proses pembersihan teks selesai dilakukan, langkah selanjutnya adalah melakukan perhitungan bobot menggunakan TF-IDF. Perhitungan dengan cara ini memberikan bobot relasi pada tiap suku kata terhadap kata. Kata tunggal pada setiap kalimat akan dianggap sebagai dokumen. Proses perhitungan TF-IDF dapat dilihat pada tabel 6 sebagai berikut.

Tabel 6. Tabel Perhitungan TF-IDF

Token	Term Frequency (Tf)					Df	D/Df	IDF	W= Tf * IDF				
	D1	D2	D3	D4	D5				W _{D1}	W _{D2}	W _{D3}	W _{D4}	W _{D5}
adalah	1	0	0	0	0	1	5	0,69	0,69	0	0	0	0
akhir	0	0	0	1	0	1	5	0,69	0	0	0	0,69	0
asli	0	0	0	1	0	1	5	0,69	0	0	0	0,69	0
bagus	0	1	0	0	0	1	5	0,69	0	0,69	0	0	0
banget	0	0	0	1	0	1	5	0,69	0	0	0	0,69	0
bener	0	2	0	0	0	1	5	0,69	0	1,39	0	0	0
bikin	0	0	0	0	1	1	5	0,69	0	0	0	0	0,69
buat	0	1	0	0	0	1	5	0,69	0	0,69	0	0	0
tonton	1	0	0	0	0	1	5	0,69	0,69	0	0	0	0
film	1	1	0	1	1	4	1,25	0,09	0,09	0,09	0	0,09	0,09
gua	0	1	0	0	1	2	2,5	0,39	0	0,39	0	0	0,39
...
yang	0	1	0	0	0	1	5	0,69	0	0,69	0	0	0

3.1.3 K-Nearest Neighbor

Dalam proses analisis sentimen menggunakan metode K-Nearest Neighbor, proses selanjutnya setelah mencari bobot setiap token dalam dokumen adalah menghitung jarak dari data training ke data uji. Pada proses ini, peneliti menggunakan rumus Euclidean Distance sebagai perhitungan jarak. Dokumen yang memiliki jarak terkecil akan mewakili kelas sentimen data uji.

Penyelesaian :

$$\begin{aligned}
 D1 &= \sqrt{\frac{(0 - 0,6989)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2}{(0 - 0)^2 + (0,6989 - 0)^2 + \dots + (0 - 0)^2}} \\
 &= \sqrt{0,4866 + 0 + 0 + 0 + 0 + 0 + 0,4886 + \dots + 0} \\
 &= \sqrt{2,759507836} \\
 &= 1,661176642
 \end{aligned}$$

$$\begin{aligned}
 D2 &= \sqrt{\frac{(0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 0,6989)^2 + (0 - 0)^2 + (0 - 0)^2}{(0 - 1,3979)^2 + (0,6989 - 0)^2 + \dots + (0 - 0,6989)^2}} \\
 &= \sqrt{0 + 0 + 0 + 0,4886 + 0 + 1,9542 + 0,4886 + \dots + 0,4885} \\
 &= \sqrt{4,555387853} \\
 &= 2,134335459
 \end{aligned}$$

$$\begin{aligned}
 D3 &= \sqrt{\frac{(0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2}{(0 - 0)^2 + (0,6989 - 0)^2 + \dots + (0 - 0)^2}} \\
 &= \sqrt{0 + 0 + 0 + 0 + 0 + 0 + 0,4886 + \dots + 0} \\
 &= \sqrt{4,28153434} \\
 &= 2,069186879
 \end{aligned}$$

$$\begin{aligned}
 D4 &= \sqrt{(0-0)^2 + (0-0,4886)^2 + (0-0,4886)^2 + (0-0)^2 + (0-0,4886)^2 + (0,6989-0,4886)^2 + \dots + (0-0)^2} \\
 &= \sqrt{0 + 0,4886 + 0,4886 + 0 + 0,4886 + 0 + 0,4886 + \dots + 0} \\
 &= \sqrt{3,578269719} \\
 &= 1,891631497
 \end{aligned}$$

Tabel hasil pengukuran jarak dokumen terhadap data uji dapat dilihat pada tabel 7 sebagai berikut.

Tabel 7. Tabel Jarak Dokumen ke Data Uji

Dokumen	Tweet	Label	Jarak
Doc 1	ngeri ngeri sedap adalah film yg sangat layak ditonton	Positif	1,661177
Doc 2	film indo yang bener bener bagus buat gua itu ngeri ngeri sedap	Positif	2,134335
Doc 3	liat masku pamit mau kerja ke tempat jauh	Netral	2,069187
Doc 4	oiyaaa film ngeri ngeri sedap juga asli itu nangis jelek banget pas dr tengah ke akhir	Negatif	1,891631

3.2 Evaluasi

Pengukuran evaluasi pada penelitian ini meliputi pengukuran precision, recall, dan F1-score. Hasil perhitungan akurasi yang dilakukan terhadap 230 data akan dilakukan dengan menggunakan ratio perbandingan 6:4 (60% data training : 40% data uji). Perhitungan dengan ratio 6:4 dapat dilihat sebagai berikut.

$$\begin{aligned}
 Precision &= \frac{TP}{TP+TN+TNeu} \\
 Precision &= \frac{73}{73+17+2} \\
 Precision &= \frac{73}{92} \\
 Precision &= 0,79
 \end{aligned}$$

Berdasarkan hasil perhitungan precision diatas, maka didapat hasil persentase untuk proses precision adalah sebanyak 79%

$$\begin{aligned}
 Recall &= \frac{TP}{TP + FN + FNeu} \\
 Recall &= \frac{73}{73 + 0 + 0} \\
 Recall &= \frac{73}{73} \\
 Recall &= 1
 \end{aligned}$$

Berdasarkan hasil perhitungan dengan rumus diatas, maka diperoleh hasil persentase untuk proses recall adalah sebesar 100%.

$$F1 - Score = 2 * \frac{Precision \times Recall}{Precision + Recall}$$

$$F1 - Score = 2 * \frac{0,79 \times 1}{0,79 + 1}$$

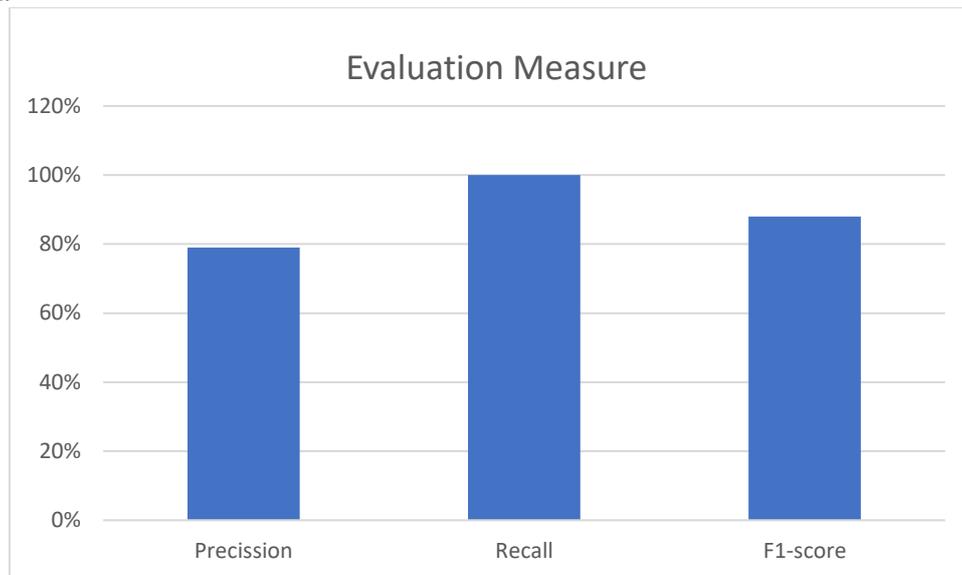
$$F1 - Score = 2 * \frac{0,79}{1,79}$$

$$F1 - Score = 2 * 0,44$$

$$F1 - Score = 0,88$$

Berdasarkan perhitungan dengan menggunakan rumus F1-score tersebut, maka dihasilkan persentase nilai F1-score sebesar 88%.

Berikut tampilan grafik berupa barchart dari proses evaluation measure pada sistem sesuai dengan perhitungan yang telah dilakukan. Dapat dilihat pada gambar 4.8 dibawah ini.



Gambar 2. Barchart Evaluation Measure

4. KESIMPULAN

Berdasarkan hasil penelitian analisis sentimen yang dilakukan pada penelitian ini, maka ditarik kesimpulan bahwa Metode K-Nearest Neighbor (KNN) mampu menggolongkan sentimen baru kedalam sentimen positif, sentimen negatif, dan sentimen netral berdasarkan data training yang telah tersedia dengan melalui pembobotan TF-IDF dan perhitungan jarak Euclidean Distance. Dari 92 data uji yang terdiri dari 73 data kelas positif, 17 data kelas netral, dan 2 data kelas negatif, sistem berhasil memprediksi 73 data kelas positif, 0 data kelas netral, dan 0 data kelas negatif. Setelah melakukan perhitungan evaluation measure pada sistem, dengan nilai k optimalnya = 5 serta menggunakan 60% data training dan 40% data testing dari 230 data set diperoleh hasil precision sebesar 79%, recall sebesar 100% dan F1-Score sebesar 88%.

REFERENSI

- [1] F. R. Irawan, A. Jazuli, and T. Khotimah, "ANALISIS SENTIMEN TERHADAP PENGGUNA GOJEK MENGGUNAKAN METODE K-NEAREST NEIGHBORS," *Jurnal Informatika dan Komputer) Akreditasi KEMENRISTEKDIKTI*, vol. 5, no. 1, 2022, doi: 10.33387/jiko.
- [2] D. Robison Manalu, M. Christofell, L. Tobing, and M. Yohanna, "METHOMIKA: Jurnal Manajemen Informatika & Komputerisasi Akuntansi ANALISIS SENTIMEN TWITTER TERHADAP WACANA PENUNDAAN PEMILU DENGAN METODE SUPPORT VECTOR MACHINE," vol. 6, no. 2, 2022, doi: 10.46880/jmika.Vol6No2.pp149-156.

-
- [3] D. P. Utomo and M. Mesran, "Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 4, no. 2, p. 437, Apr. 2020, doi: 10.30865/mib.v4i2.2080.
- [4] O. Sakti, Y. Prakasa, and K. M. Lhaksamana, "KLASIFIKASI TEKS DENGAN MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR PADA KASUS KINERJA PEMERINTAH DI TWITTER," 2018.
- [5] M. I. Hutapea and A. P. Silalahi, "Moderna's Vaccine Using the K-Nearest Neighbor (KNN) Method: An Analysis of Community Sentiment on Twitter," *Jurnal Penelitian Pendidikan IPA*, vol. 9, no. 5, pp. 3808–3814, May 2023, doi: 10.29303/jppipa.v9i5.3203.
- [6] M. A. Rofiqi, Abd. C. Fauzan, A. P. Agustin, and A. A. Saputra, "Implementasi Term-Frequency Inverse Document Frequency (TF-IDF) Untuk Mencari Relevansi Dokumen Berdasarkan Query," *ILKOMNIKA: Journal of Computer Science and Applied Informatics*, vol. 1, no. 2, pp. 58–64, Dec. 2019, doi: 10.28926/ilkomnika.v1i2.18.
- [7] A. P. Silalahi and H. G. Simanullang, "SUPERVISED LEARNING METODE K-NEAREST NEIGHBOR UNTUK PREDIKSI DIABETES PADA WANITA," *METHOMIKA Jurnal Manajemen Informatika dan Komputerisasi Akuntansi*, vol. 7, no. 1, pp. 144–149, Apr. 2023, doi: 10.46880/jmika.Vol7No1.pp144-149.
- [8] J. B. Lubis *et al.*, "Penerapan Metode K-Nearest Neighbor Dalam Mengklasifikasi Status Ekonomi Suatu Penduduk," 2023. [Online]. Available: <http://ojs.fikom-methodist.net/index.php/methosisfo>
- [9] R. Perangin-Angin, E. Julia, G. Harianja, and I. K. Jaya, "PENDEKATAN LEVEL DATA UNTUK MENANGANI KETIDAKSEIMBANGAN DATA MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR," 2020. [Online]. Available: <https://sci2s.ugr.es/keel/datasets.php>
- [10] R. Ramadhan Al-Mubaraq, S. Al Faraby, and M. Dwifabri Purbolaksono, "Analisis Sentimen pada Ulasan Film dengan Kombinasi Seleksi Fitur Chi-Square dan TF-IDF menggunakan Metode KNN," 2021.
- [11] H. Jurnal and R. Subariah, "JURNAL PUBLIKASI TEKNIK INFORMATIKA PERANCANGAN APLIKASI PERSEDIAAN BAHAN KUE BERBASIS WEBSITE MENGGUNAKAN METODE PROTOTYPE PADA DAPUR ALLYSA," *JUPTI*, vol. 2, no. 2, 2023.