

Klasifikasi Sentimen Positif Dan Negatif Pada Aplikasi Vidio Dengan Algoritma K-Nearest Neighbor

Muhammad Fudhail Ferio Supeli¹, Setiaji²

Universitas Nusa Mandiri^{1 2},

mfudhail972015@gmail.com¹, setiaji.sej@nusamandiri.ac.id²

Diterima (09-03-2023)	Direvisi (26-03-2023)	Disetujui (18-04-2023)
--------------------------	--------------------------	---------------------------

Abstrak - Keadaan pada tahun 2022 sekarang masih sedikit diselimuti wabah virus *covid-19*. Kondisi sekarang mungkin wabah *covid-19* mulai menurun, tetapi dengan kebijakan yang dikeluarkan oleh pemerintah sebelumnya membuat beberapa dari masyarakat peminat *film* dan pertandingan olahraga terlanjur terbiasa di rumah. *Vidio* adalah sebuah media atau aplikasi yang menyediakan layanan *streaming online*, berupa *film*, *drama*, *series*, dan ada juga *live streaming* pertandingan olahraga. Aplikasi ini sekarang sedang banyak diminati pelanggan karena fitur-fitur atau promo yang dikeluarkan sangat menarik. Berhubungan dengan hal tersebut, penelitian terfokus kepada sentimen ulasan pengguna aplikasi *vidio* di *google play store*. Penelitian ini bertujuan untuk mengetahui klasifikasi positif dan negatif pada ulasan tersebut, dengan harapan aplikasi *vidio* bisa berkembang lebih bagus, terlihat sekarang aplikasi *Vidio* berkembang pesat tetapi masih banyak keluhan yang ditemukan didalam ulasannya. Tujuan penelitian ini diantara lain yaitu mengimplementasikan analisis sentimen dengan menggunakan metode *K-Nearest Neighbor* dan juga mengklasifikasikan kalimat sentimen pada ulasan ke dalam kelompok positif dan negatif pada aplikasi *Vidio*. Metode yang digunakan dalam penelitian ini yaitu algoritma *K-Nearest Neighbor* dengan dua cara perhitungan yaitu manual dan menggunakan tools *rapidminer*. Hasil yang didapat berdasarkan perhitungan manual yaitu mendapatkan 70% akurasi. Sedangkan dengan menggunakan *rapidminer* mendapatkan 50% akurasi.

Kata Kunci : *Vidio*, *K-Nearest Neighbor*, *Rapidminer*

Abstract - The situation in 2022 is still slightly shrouded in the *covid-19* virus outbreak. The current condition may be that the *COVID-19* outbreak is starting to decline, but with the policies issued by the previous government, some of the people who are interested in films and sports competitions are already accustomed to staying at home. *Vidio* is a media or application that provides online streaming services, in the form of films, dramas, series, and there is also live streaming of sports matches. This application is now in great demand by customers because the features or promos issued are very attractive. In this regard, the research focuses on the sentiment of user reviews of *Vidio* applications on the *Google Play Store*. This study aims to determine the positive and negative classifications in the review, with the hope that the video application can develop better, it can be seen now that the *Vidio* application is growing rapidly but there are still many complaints found in the review. The objectives of this study include implementing sentiment analysis using the *K-Nearest Neighbor* method and also classifying sentiment sentences in reviews into positive and negative groups in the *Vidio* application. The method used in this research is the *K-Nearest Neighbor* algorithm with two calculation methods, namely manual and using *rapidminer* tools. The results obtained are based on manual calculations, get 70% accuracy. Meanwhile, by using *rapidminer* get 50% accuracy.

Keywords: *Vidio*, *K-Nearest Neighbor*, *Rapidminer*

I. PENDAHULUAN

Wabah virus *covid-19* sudah menyerang indonesia terhitung kurang lebih dua tahun. banyak kebijakan yang dikeluarkan pemerintah indonesia untuk memutus mata rantai penyebaran *covid-19* di indonesia. Seperti PSBB (pembatasan sosial berskala besar), lalu PPKM (pemberlakuan pembatasan kegiatan

masyarakat) yang mengharuskan seluruh masyarakat untuk melakukan aktivitas di rumah saja.

Walaupun saat ini kebijakan pemerintah berhasil mengurangi penyebaran *covid-19* di tahun 2022 ini, dan banyak tempat yang sudah dibuka, masyarakat masih banyak yang terlanjur terbiasa menonton di rumah. karena sebelumnya masyarakat tidak bisa ke bioskop

untuk menonton *film*, dan tidak bisa menonton pertandingan olahraga secara langsung maka bisa dipastikan mereka yang hobi menonton mendownload aplikasi *mobile* penyedia layanan *streaming film* dan *sports* di *smartphone* dalam kondisi sebelumnya sampai sekarang. ada kemungkinan, pecinta film hilang minat untuk ke bioskop karena sekarang sudah bisa menonton film dan pertandingan olahraga di rumah dengan aplikasi *mobile* layanan *streaming film* salah satunya yaitu aplikasi *Vidio*.

Vidio, atau *Vidio.com* adalah aplikasi hiburan penyedia *streaming film* indonesia dan film luar negeri, dan juga salah satu *platform streaming* pertandingan olahraga live. untuk *streaming*, tidak bisa semua dinikmati secara penuh karena harus berlangganan *premier* terlebih dahulu. Rating aplikasi ini cukup untuk dibidang lumayan bagus yaitu 4.2 bintang berdasarkan dari ulasan aplikasi *vidio* yang tertera di *google play store* pada tahun 2022. *google* memiliki sebuah layanan yang dikenal dengan nama *playstore* yang menyediakan konten-konten digital seperti game, aplikasi, film, musik, dan buku dengan kategori yang beragam (Agustina C.A et al., 2022). contohnya seperti aplikasi *vidio* ini yang bisa kita download secara gratis. aplikasi ini mendapat predikat nomor 1 terlaris di kategori hiburan.

Walaupun sekarang aktivitas sudah tidak dibatasi seperti saat pandemi, masih banyak orang yang menikmati *streaming film* atau *streaming pertandingan olahraga* di rumah lewat *smartphone* salah satunya dengan berlangganan aplikasi *vidio* pasti pelanggan memberikan banyak komentar. Suatu aplikasi selalu memiliki kekurangan dan kelebihan masing-masing, dimana hal tersebut dapat menimbulkan berbagai respon dari pengguna aplikasi seperti kepuasan dan kekecewaan terhadap aplikasi tersebut (Giovani et al., 2020) Dari banyak komentar itu ada yang berisi positif dan negatif terhadap aplikasi tersebut. Komentar dari pelanggan yang menggunakan aplikasi *vidio* sangat berguna untuk evaluasi bagi pembuat aplikasi *vidio* untuk melihat beberapa kekurangan dan memperbaikinya. Dalam menganalisis sentiment pada aplikasi *vidio*, penulis menggunakan metode algoritma *K-Nearest Neighbor*. *K-Nearest Neighbor* adalah salah satu metode dimana metode ini melakukan klasifikasi berdasarkan data training atau data pembelajaran dilihat dari jarak yang paling dekat dengan objek berdasarkan nilai *K* (Ambar S et al., 2018)

Analisis sentimen atau penambangan opini adalah suatu bidang studi untuk menganalisis pendapat orang terhadap entitas seperti produk, layanan, organisasi, individu, masalah,

peristiwa, dan topik (Septian et al., 2019).

Dari penjelasan latar belakang diatas penulis akan melakukan analisis sentimen positif dan negatif pada aplikasi *vidio* menggunakan algoritma *K-Nearest Neighbor*, untuk mendapatkan hasil tertinggi Identifikasi masalah dari penelitian ini adalah

1. Mengimplementasikan metode algoritma *K-Nearest Neighbor (K-NN)* untuk mendapatkan hasil pengujian yang akurat dalam mengklasifikasi komentar positif maupun negatif pada aplikasi *vidio*
2. Setelah masyarakat terpaksa harus beraktivitas di rumah, mungkin masih banyak yang menggunakan aplikasi *mobile* penyedia *streaming film* dan pertandingan olahraga karena kekecewaan sebelumnya yang tidak bisa menonton secara langsung ke bioskop atau stadion semenjak adanya kebijakan di rumah saja dari pemerintah. bisa dibidang sebagian masyarakat terlanjur menikmati *streaming* di rumah lewat aplikasi *mobile*.

Tujuan pada penelitian ini yaitu :

1. Mengimplementasikan analisis sentiment dengan menggunakan metode *K-Nearest Neighbor*.
2. Mengklasifikasikan kalimat sentiment pada ulasan ke dalam kelompok positif dan negatif pada aplikasi *vidio*.

Berikut beberapa penelitian sebelumnya yang terkait dengan penelitian ini. Penelitian Retno (Sari, 2020) membahas tentang "Analisis Sentimen Pada Review Objek Wisata Dunia Fantasi menggunakan Algoritma *K-Nearest Neighbor (K-NN)*", hasil penelitian tersebut menunjukkan akurasi sebesar 77,01% menggunakan metode algoritma *K-Nearest Neighbor*. Selain itu, penelitian Abu, Junta, & Rima (Salam et al., 2018) membahas tentang "Analisis Sentimen Data Komentar Sosial Media Facebook Dengan *K-Nearest Neighbor* (Studi Kasus Pada Akun Jasa Ekspedisi Barang J&T Ekspres Indonesia)", hasil penelitian tersebut menunjukkan akurasi tertinggi sebesar 79,21% menggunakan metode algoritma *K-Nearest Neighbor*. Sedangkan penelitian Rifki & Anggi (Kosasih & Alberto, 2021) yang membahas tentang "Analisis Sentimen Produk Permainan Menggunakan Metode TF-IDF Dan Algoritma *K-Nearest Neighbor*", dalam penelitian tersebut menghasilkan akurasi sebesar 79,3333%. dari tiga penelitian terkait yang telah disebutkan

hasilnya diatas, memiliki tujuan untuk mendukung permasalahan yang terjadi pada penelitian ini.

II. METODOLOGI PENELITIAN

Tahapan pada penelitian ini dideskripsikan sebagai berikut :

1. Identifikasi Masalah
Langkah pertama yang dilakukan pada penelitian ini yaitu identifikasi masalah yang bertujuan agar hasilnya tidak keluar dari alur penelitian.
2. Tujuan Penelitian
Tujuan penelitian ini adalah untuk mendapatkan hasil akurasi yang terbaik dari klasifikasi komentar positif dan negatif pada aplikasi *vidio*.
3. Tinjauan Pustaka
Di dalam tahapan tinjauan pustaka bertujuan untuk mengumpulkan teori-teori dari para ahli dan pakar yang berfungsi untuk mendukung teori dalam penelitian ini, agar penelitian ini bersifat logis dan terarah karena mengandung teoridari para ahli dan pakar.
4. Penentuan Sampel
Dalam tahapan ini, penentuan sampel yang digunakan yaitu ulasan atau komentar pengguna aplikasi *Vidio* pada bulan Maret sampai Juni 2022.
5. Pengumpulan dan Pengolahan Data
Data-data yang digunakan dalam penelitian ini adalah data berupa ulasan atau komentar dari pengguna aplikasi *Vidio*, di dalam *Google Playstore*. Pengumpulan data menggunakan Teknik scrapping yang dimana data tidak terstruktur ditambang dari ulasan aplikasi *Vidio* tersebut. Lalu data tersebut dilabelkan menjadi dua kategori positif dan kategori negatif.
6. Analisis Data
Dalam tahapan analisis data, yang dilakukan adalah tahapan pelabelan manual yang dimana pelabelan untuk membagi data menjadi dua kategori yaitu kategori positif yang berisi ulasan berupa apresiasi atau pujian dan dukungan sedangkan kategori negatif berupa ulasan yang menggambarkan hinaan, protes dan yang lainnya. Setelah itu dilakukannya tahap *preprocessing* yang didalamnya data akan melewati teknik yang berupa :

- a. Case Foulding
Merupakan proses mengubah huruf menjadi huruf kecil, atau disebut juga *lowercase*.
- b. Cleansing
Adalah proses menghapus semua karakter yang tidak termasuk alfabet, untuk mengurangi *noise* pada dataset. Contoh karakter yang dihapus yaitu jika ada tanda @, #, (, .) dan berbagai variasi tanda baca lainnya.
- c. *Tokenizing*
Adalah proses yang dilakukan untuk memotong kalimat menjadi berbagai kata.
- d. *Stopword*
Adalah tahapan membuang kata-kata yang tidak bermakna, memperbaiki salah ketik misalnya data yang tidak konsisten di proses ini akan diperbaiki.
- e. *Stemming*
Adalah proses untuk mengubah kata, dari kata imbuhan menjadi kata dasar.

Setelah melewati *preprocessing* dan tahapannya, dilakukannya tahap pengujian dengan algoritma *K-Nearest Neighbour* menggunakan *Microsoft Excel* dan aplikasi *Rapidminer* dan menentukan hasil akurasinya.

7. Kesimpulan dan Saran

Tahap yang terakhir yaitu kesimpulan dan saran, yang berisi kesimpulan dari penelitian yang dilakukan dan saran yang diberikan penulis untuk usulan pengembangan pada pengelola aplikasi *Vidio*.

Pengumpulan data dilakukan dengan menganalisis kumpulan data yang berupa ulasan aplikasi *vidio* di *Google Play Store* yang tersedia dalam bentuk data yang tidak terstruktur. pengumpulan data juga dilakukan dengan cara studi literatur, yaitu dengan mengumpulkan jurnal, dan buku-buku yang berkaitan dengan judul penelitian
Sampel data diambil dari 500 data dengan rumus *slovin*(1) sebagai berikut :

$$n = \frac{N}{1+N \cdot e^2} \dots\dots\dots(1)$$

n = Jumlah Sampel
N = Jumlah Populasi
e = Taraf keyakinan peneliti mentolerir tingkat kesalahan sebesar 0,05 menjadi:

$$n = \frac{500}{1+500 \cdot (0,05)^2}$$

$$= 222,222 \approx 222$$

Maka dapat disimpulkan sampel yang akan digunakan berjumlah 222 buah data sampel.

III. HASIL DAN PEMBAHASAN

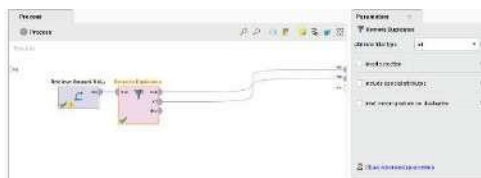
Ulasan pada aplikasi *Vidio* yang masih tidak terstruktur sebanyak 500 buah data yang akan melalui tahapan-tahapan agar data tersebut menjadi terstruktur dan siap digunakan. Berikut adalah tahapan-tahapan dalam Analisa data pada penelitian ini, yaitu :

Dalam penelitian ini, pengumpulan data menggunakan metode *scrapping* dengan tools yang digunakan yaitu *Webharvy*, ulasan yang dikumpulkan adalah ulasan dengan bentuk yang tidak terstruktur sebanyak 500 data.



Sumber : Peneliti (2022)
Gambar 1. *Scrapping* data dengan *Webharvy*

Data yang telah selesai di *Scrapping* akan di *export* dalam bentuk *Excel*, selanjutnya yaitu menuju ke tools *Rapidminer* untuk melalui proses *remove duplicates* yaitu tahapan untuk menghilangkan duplikat pada data sebelum menuju proses pelabelan dan *preprocessing*.



Sumber : Peneliti (2022)
Gambar 2. *Remove Duplicates*

Setelah melalui proses *remove duplicates*, dan sebelum melalui proses *preprocessing*, data siap untuk dilabelkan secara manual, yang

bertujuan untuk mengubah data dari tidak terstruktur menjadi data yang terstruktur pelabelan akan dibagi menjadi dua kelas yaitu kelas positif dan kelas negatif. untuk membaginya, data yang terdapat 5,4,3 bintang diberi label positif dan yang terdapat 2,1 bintang diberi label negatif.

Tabel 1. Contoh dataset

Nama	Tanggal	Rating	Ulasan
Jakson	3 Juni	Diberi 5	Untuk semua
Simanjun	2022	Dari 5	film bagus
		Bintang	sekali.
			Terlebih film-
			film yang ada
			di bioskop
			banyak
			tersedia di
			apk vidio.
			suka-suka
			sekali nonton
			di Vidio.
			Semoga bug
			pada
			layar saat di
			zoom
			diperbaiki.
			Serta lebih
			ditingkatk an
			lagi
			pengatura n
			gaya
			subtitle dan
			ukuran
			subtitle agar
			lebih
			nyaman
			Ketika
			menonton.
			Sukses terus
			APK
			vidio 🤔👍

Dan berikut tabel dari contoh data yang telah dilabelkan secara manual :

Tabel 2. Data yang telah dilabelkan secara manual

Ulasan	Sentimen
Ingin sekali	Negatif
berlangganan lagi	
tapi sangat sulitdalam	
melakukan	
pembayaran, tidak seperti	
pembayaran vidio yang	
dulu dapat	
dilakukan denganmudah	
lewataplikasi	
Dana dll. Pembayaran	
sangatrumit,	
mau	
berlangganan jadimales.	

Untuk semua film bagus Positif sekali.

Terlebih film-film yang ada di bioskop banyak tersedia di apk vidio. suka-suka sekali nonton di Vidio. Semoga bug pada layar saat di zoom diperbaiki. Serta lebih ditingkatkan lagi pengaturan gaya subtitle dan ukuran subtitle agar lebih nyaman

ketika menonton. Sukses terus
APK

vidio 🤗👍

Preprocessing merupakan tahapan awal dalam mengolah data input sebelum memasuki proses utama (Gunawan Sudarsono et al., 2021). dimana data dipersiapkan menjadi data yang nantinya akan dianalisis pada proses selanjutnya. Tahapan ini terdiri dari proses *case foulding*, *cleansing*, *tokenizing*, *stopwords*, dan *stemming*.

1. Case Foulding

Case Foulding merupakan tahapan merubah semua huruf yang ada di data menjadi huruf kecil atau lower case, berikut gambar dari tahapan *case foulding* :



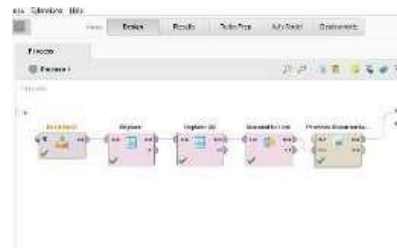
Sumber : Peneliti (2022)

Gambar 3. Tahapan *Case Foulding*

2. Cleansing

Tahapan *cleansing* adalah tahapan menghapus semua karakter yang tidak termasuk alfabet, untuk mengurangi *noise* pada dataset. Contoh karakter yang dihapus yaitu jika ada tanda @, #, (,) dan berbagai variasi tanda baca lainnya.

Berikut contoh gambar dari tahapan *cleansing* :



Sumber: Peneliti (2022)

Gambar 4. Tahapan *Cleansing*

c. Tokenizing

Adalah proses yang dilakukan untuk memotong kalimat menjadi berbagai kata, menghilangkan tanda baca yang tidak diperlukan dalam data. Berikut contoh gambar dari tahapan *Tokenizing* :

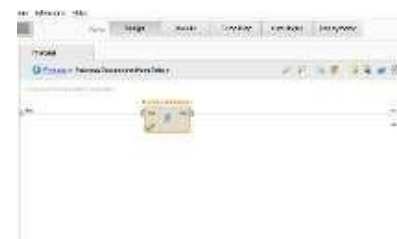


Sumber : Peneliti (2022)

Gambar 5. Tahapan *Tokenizing*

d. Stopword

Adalah tahapan membuang kata-kata yang tidak bermakna, memperbaiki salah ketik misalnya data yang tidak konsisten di proses ini akan diperbaiki. Berikut contoh gambar dari tahapan *Stopword* :



Sumber : Peneliti (2022)

Gambar 6. Tahapan *Stopword*

e. Stemming

Adalah proses untuk mengubah kata, dari kata imbuhan menjadi kata dasar, memperbaiki bahasa.

Berikut contoh gambar dari tahapan *Stemming* :



Sumber : Peneliti (2022)

Gambar 7. Tahapan *Stemming*

Berikut ini contoh data yang telah melewati proses preprocessing:

Tabel 3. Data yang telah melewati proses preprocessing

Sebelum	Sesudah
harga	harga
langganan	langganan
terjangkau	terjangkau
langganan	langganan
final uecl ucl	final uecl ucl
request	request
konten agenda	konten agenda
timnas level u	timnas level u
senior adakan	senior adakan
sports	sports
memiliki	memiliki
menu	menu sendiri
tersendiri	overall oke
overall okebanget	banget sih
sih	

Selanjutnya adalah tahapan pembobotan TF-IDF yang bertujuan untuk memberikan suatu nilai bobot pada suatu ulasan yang akan dijadikan sebagai data testing pada algoritma *K-Nearest Neighbor*. Ada 10 dataset yang akan diambil dengan kata kunci bagus yang dipilih. Yaitu: nonton, iklan, aplikasi, paket, pembayaran, bagus, film, berlangganan, beli.

Tabel 4. 10 dataset

D1	pembayaran berlangganan uang google play mengirim email terang beli akun vidio paket aktif tulisan nya pembayaran diproses menit nonton bola ketinggalan menyebalkan	Negatif
D2	suka banget apk nonton film ketinggalan tapi iklannya suka loading terus tolong	Negatif

	perbaiki kalo kaya gini orang kesel tolong perbaiki biar nyaman kebalinonton	
D3	langganan premium tayangan motogp nya batasi nya buka akses tv lokal gak sesuai iklan nya not recommended	Negatif
D4	aplikasinya bagus iklannya ruamayan mengganggu pas nonton seru ajatapitetep bagus	Positif
D5	udah bagus bagus banget paket premier platinumsesuai lh semoga kedepan nya pertahankan harga nya sesuai namanya gak yg gratis klo gratis bikin siaran	Positif
D6	aplikasi bagus ku nonton tapi mohon kurangin iklan nya nonton iklan nya trus iklan nya mohon perbaiki maaf ksh bintang	Negatif
D7	gapernah ngulas aplikasi kali benar benar kesel pas dah film pas seru nya muncul koneksi error koneksi bagus	Negatif
D8	sebener nyah app bagus iklan nyah bayak kadang kadang iklan yg ga pencet skip iklan bayak	Positif
D9	bayak minimal detik detik x iklan nyah aplikasi nya bagus kualitas video bagus ubah film nya lengkap mohon dilengkapi ya semoga aplikasinya bagus depan	Positif
D10	aplikasinya bagus bantu konten nya menarik gimna atasi vidio offline ga putarharus paket data diputar ya aja kaya ga download vidio offline mohon bantuannya	Positif

Dari 10 dataset yang ditentukan, selanjutnya adalah proses pembobotan dengan algoritma TF-IDF. Metode TF-IDF merupakan metode pembobotan *term* secara statistik yang banyak digunakan sebagai vektorisasi pada teks analisis (Baskoro et al., 2021).Berikut adalah gambar dari pembobotan algoritma TF-IDF :

Sumber : Peneliti (2022)

Gambar 8. Pembobotan TF-IDF

Setelah melakukan pembobotan TF-IDF, bisa terlihat hasilnya yang disebut ranking, untuk melakukan perankingan. Tetapi di perhitungan TF-IDF tersebut ada hasil dari ranking yang menunjukkan angka kembar di D7 dan D10. Maka yang harus dilakukan adalah mencari lagi perankingan dengan menggunakan algoritma *Vektor Space Model*, menghitung jarak bobot (*weighted distance*) dengan dipangkatkan 2. Berikut adalah perhitungan dengan algoritma *Vektor Space Model* :

Sumber : Peneliti (2022)

Gambar 9. Algoritma Vektor Space Model

Setelah melakukan perhitungan algoritma Vektor Space Model, lalu ditemukanlah hasil dari perankingan yang tidak kembar dari data ke 1 sampai data ke 10 yang sudah juga diurutkan datanya.

Berikut gambar perankingan TF-IDF yang selesai dihitung

D1 KE D1 - 10		
0,58940	10	POSITIF
0,51551	5	POSITIF
0,48909	8	POSITIF
0,48300	7	NEGATIF
0,44929	9	POSITIF
0,44311	2	NEGATIF
0,38741	6	NEGATIF
0,33053	4	POSITIF
0,22805	3	NEGATIF
0	1	NEGATIF

Sumber : Peneliti (2022)

Gambar 10. Hasil Perankingan TF-IDF

Setelah selesai melakukan perankingan dengan tahap algoritma TF-IDF, lalu dihitunglah jarak tiap data. Perhitungan menggunakan rumus dari algoritma *K-Nearest Neighbor* yaitu rumus *euclidean distance*, yang merupakan cara yang sering digunakan untuk menghitung jarak antar data (Nikmatul Kasanah et al., 2019).

Berikut rumus dari *Euclidean distance*(2)

$$= \sqrt{\sum_{i=1}^n (q_i - p_i)^2} \dots \dots \dots (2)$$

- i = setiap data
- n = jumlah data,
- qi = data pada pusat klaster ke ipi = data pada setiap data ke i

Berikut gambar dari contoh perhitungan Euclidian Distance :

D1	0,6761212
D3	0,4480758
D4	0,3455863
D6	0,2887043
D2	0,2330153
D9	0,2268275
D7	0,1931195
D8	0,1870301
D5	0,1606032
D10	0,0867176

Sumber : Peneliti (2022)

Gambar 11. Hasil Perhitungan Euclidian Distance

Lakukan sampai D1 ke D10-10 dengan K =7 Setelah menghitung dengan rumus *euclidean distance*, maka selanjutnya adalah tahap mencari akurasi(3) dengan *confusion matrix* dengan rumus sebagai berikut :

$$= \frac{TP+TN}{TP+TN+FP+FN} \dots \dots \dots (3)$$

- TP = True Positive
- TN = True Negative
- FP = False Positive
- FN = False Negative

Berikut contoh gambar dari Confusion Matrix :

		True Class	
		Positive	Negative
Predicted Class	Positive	TP	FP
	Negative	FN	TN

Sumber : towardsdatascience.com

Gambar 12. Confusion Matrix

Dan berikut contoh perhitungan manual Confusion Matrix

	1	AKTUAL	
	PREDIKSI	POSITIF	NEGATIF
POSITIF		3	1
NEGATIF		2	4

Sumber : Peneliti (2022)

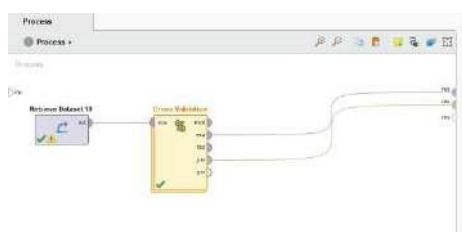
Gambar 13. Hasil Perhitungan Confusion Matrix

Lakukan perhitungan sampai ke-10, lalu carilah akurasi dengan menggunakan rumus mencari akurasi confusion matrix sesuai dengan rumusnya, dan hasil yang didapat dari perhitungan ini 0,7 per hasil confusion matrix lalu di cari rata-ratanya dengan seluruh hasil 1-10, mendapat 70%.

Dari perhitungan 10 data diatas, berhasil ditemukanlah akurasi dengan K=7 Sebesar 70% menggunakan perhitungan manual *Microsoft Excel* Sekarang lalu kita mencoba perhitungan dengan menggunakan *Rapidminer*.

Data perhitungan yang digunakan untuk menghitung klasifikasi dengan *rapidminer* yaitu dataset yang telah melewati, preprocessing. Data berjumlah 10 dipilih dengan 5 kelas positif dan 5 kelas negatif. Data dihitung dengan *10-fold cross validation* menggunakan *rapidminer*.

Berikut gambar perhitungan Rapidminer :



Sumber : Peneliti (2022)

Gambar 14. Perhitungan Rapidminer

Lalu, didapatkan hasil akurasi seperti gambar dibawah ini, berikut adalah gambar dari hasil akurasi yang didapat :

accuracy: 80.00% +/- 52.70% (size average: 5000%)

	Real NEGATIF	Real POSITIF	100% precision
Real NEGATIF	5	0	100%
Real POSITIF	0	5	100%
class recall	100.00%	100%	

Sumber : Peneliti (2022)

Gambar 15. Hasil Akurasi Rapidminer

Dengan k=7, didapatkan akurasi sebesar 50.00% dari perhitungan menggunakan *rapidminer*, dari 10 data ulasan.

IV. KESIMPULAN

Berdasarkan hasil penelitian dan hasil analisis yang didapatkan, tentang klasifikasi sentiment positif dan negatif pada aplikasi vidio dengan algoritma *K- Nearest Neighbor* dapat disimpulkan sebagai berikut dibawah ini.

Dari hasil uji yang dilakukan, didapat akurasi dari perhitungan manual menggunakan *Microsoft Excel* sebesar 70%. Dan akurasi dari perhitungan menggunakan *Rapidminer* sebesar 50%. Hasil dari uji yang dilakukan pada *Microsoft Excel* lebih baik daripada menggunakan tools *Rapidminer*.

Kelebihan dari algoritma K-Nearest Neighbor yaitu sangat mudah diimplementasikan apalagi dengan perhitungan secara manual. Akan tetapi jika datanya berjumlah sangat banyak akan sangat sulit untuk dihitung. Itulah yang menjadi sebuah kekurangannya.

V. REFERENSI

Agustina C.A, N., Herlina Citra, D., Purnama, W., Nisa, C., & Rozi Kurnia, A. (2022). Implementasi Algoritma Naive Bayes untuk Analisis Sentimen Ulasan Shopee pada Google Play Store. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 2(1), 47–54.

Ambar S, Y., Kusri, & Henderi. (2018). Penerapan Algoritma K-Nearest Neighbour Dalam Menentukan Pembinaan Koperasi Kabupaten Kotawaringin Timur. *Universitas AMIKOM Yogyakarta*, 5(3), 232–241.

Baskoro, B. B., Susanto, I., & Khomsah, S. (2021). Analisis Sentimen Pelanggan Hotel di Purwokerto Menggunakan Metode Random Forest dan TF-IDF (Studi Kasus: Ulasan Pelanggan Pada Situs TRIPADVISOR). *Journal of Informatics, Information System, Software Engineering and Applications*, 3(2), 22–29.

Giovani, A. P., Ardiansyah, Haryanti, T., Kurniawati, L., & Gata, W. (2020). Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi. *Jurnal Teknoinfo*, 14(2), 115. <https://doi.org/10.33365/jti.v14i2.679>

Gunawan Sudarsono, B., Ignatius Leo, M., Santoso, A., & Hendrawan, F. (2021).

- ANALISIS DATA MINING DATA NETFLIX MENGGUNAKAN APLIKASI RAPID MINER. *Journal of Business and Audit Information Systems*, 4(1), 13–21.
- Kosasih, R., & Alberto, A. (2021). Analisis Sentimen Produk Permainan Menggunakan Metode TF-IDF Dan Algoritma K-Nearest Neighbor. *InfoTekJar: Jurnal Nasional Informatika Dan Teknologi Jaringan*, 6(1), 135–139.
- Nikmatul Kasanah, A., Muladi, & Pujiyanto, U. (2019). Penerapan Teknik SMOTE untuk Mengatasi Imbalance Class dalam Klasifikasi Objektivitas Berita Online Menggunakan Algoritma KNN. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 3(2).
- Salam, A., Zeniarja, J., & Khasanah, R. S. U. (2018). ANALISIS SENTIMEN DATA KOMENTAR SOSIAL MEDIA FACEBOOK DENGAN K-NEAREST NEIGHBOR (STUDI KASUS PADA AKUN JASA EKSPEDISI BARANG J&T EKSPRESS INDONESIA). *Prosiding SINTAK 2018*, 2.
- Sari, R. (2020). Analisis Sentimen Pada Review Objek Wisata Dunia Fantasi menggunakan Algoritma K-Nearest Neighbor (K-NN). *Evolusi: Jurnal Sains Dan Manajemen*, 8(1), 10–17.
- Septian, J. A., Fahrudin, T. M., & Nugroho, A. (2019). Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-Nearest Neighbor. *JOURNAL OF INTELLIGENT SYSTEMS AND COMPUTATION*, 1(1), 43–49.