OKTAL : Jurnal Ilmu Komputer dan Science
Volume 3, No. 4, April 2024
ISSN 2828-2442 (media online)
Hal 865-874

Analisis Sentimen Terhadap Industri *E-Sports* Pada Media Sosial *Twitter* Dengan Menggunakan Metode *K-Nearest Neighbor*

Desi Jasmiati[1*], Hidayatullah Al Islami[1]

[1]Fakultas Ilmu Komputer, Teknik Informatika, Universitas Pamulang, Jl. Raya Puspiptek No. 46, Kel. Buaran, Kec. Serpong, Kota Tangerang Selatan. Banten 15310, Indonesia
Email: [1*]dessyjasmiati@gmail.com, [2]dosen02408@unpam.ac.id
(* : coressponding author)

Abstrak−Fenomena *e-sports* mengalami perkembangan signifikan seiring pesatnya perkembangan teknologi informasi dan komunikasi pada masa revolusi industri 4.0. Industri ini telah meraih beberapa prestasi yang menguntungkan dan menjadi topik populer yang menuai beragam komentar di media sosial, salah satunya *Twitter*. Analisis sentimen berguna untuk mengetahui kecenderungan opini masyarakat bernilai positif atau negatif, sehingga nantinya dapat menjadi informasi yang terstruktur. Data yang digunakan yakni 200 komentar *twitter*. Proses yang dilakukan yaitu *crawling, preprocessing* mencakup *cleaning*, normalisasi, *case folding, stopword removal*, *tokenizing*, dan *stemming*, melakukan pelabelan sentimen dan perhitungan bobot menggunakan *TF-IDF*. Metode klasifikasi menggunakan *K-Nearest Neighbor*, uji validasi menggunakan *10-fold cross validation* dan perhitungan akurasi memakai rumus *confussion matrix*. *Tools rapid miner* dipakai untuk memproses tahapan secara otomatis. Hasil akhir diperoleh nilai k=6 sebagai nilai tertinggi dengan akurasi sebesar 66.00%, presisi dengan nilai 67.34%, *recall* 78.06% dan AUC dengan nilai 0.762. Sedangkan hasil klasifikasi sentimen memperoleh sentimen positif lebih unggul dibanding sentimen negatif. Berdasarkan hasil tersebut, metode *K-Nearest Neighbor* dinilai dapat melakukan pengklasifikasian dengan baik dan diketahui bahwa masyarakat Indonesia saat ini menanggapi industri *e-sports* cenderung lebih memberikan stigma positif dan respon yang baik.

Kata Kunci: Analisis Sentimen, E-sports, K-Nearest Neighbor, Twitter

Abstract−*The phenomenon of e-sports has experienced significant developments along with the rapid development of information and communication technology during the industrial revolution 4.0. This industry has achieved several lucrative achievements and has become a popular topic that reaps a variety of comments on social media, one of which is Twitter. Sentiment analysis is useful for determining the tendency of public opinion to be positive or negative, so that later it can become structured information. The data used is 200 twitter comments. The processes carried out are crawling, preprocessing including cleaning, normalization, case folding, stopword removal, tokenizing, and stemming, carrying out sentiment labeling and weight calculations using TF-IDF. The classification method uses K-Nearest Neighbor, validation tests use 10-fold cross validation and accuracy calculations use the confussion matrix formula. The rapid miner tool is used to process the stages automatically. The final result obtained a value of k = 6 as the highest value with an accuracy of 66.00%, precision with a value of 67.34%, recall 78.06% and AUC with a value of 0.762. While the results of the sentiment classification obtained positive sentiment superior to negative sentiment. Based on these results, the K-Nearest Neighbor method is considered to be able to classify well and it is known that Indonesian people currently respond to the e-sports industry tends to provide more positive stigma and a good response.*

Keywords*: Sentiment Analysis, E-sports, K-Nearest Neighbor, Twitter*

1. PENDAHULUAN

Masa revolusi industri 4.0 dimana teknologi informasi dan komunikasi dapat di akses melalui internet, memungkinkan informasi bisa ditemukan dengan mudah oleh siapa saja. Selain itu pesatnya teknologi pun banyak memunculkan inovasi baru di dunia industri, salah satunya industri *game* atau *e-sports*. *E-sports* merupakan singkatan dari *electronic sports*, adalah jenis olahraga dengan penggunaan alat elektronik seperti *handphone* dan komputer. Fenomena *e-sports* mengalami perkembangan signifikan, yang membawa pengaruh besar pada industri *gaming* di Indonesia. Berdasarkan data tahun 2021, Wilayah Asia-Pasifik sejauh ini menyumbang sebagian besar pemain *game online* di seluruh dunia (55% dari jumlah global). Pada tahun 2020, Indonesia memperoleh pendapatan sebesar *USD* 1,74 miliar atau setara dengan Rp. 25 triliun dari industri *game*. Indonesia di prediksi akan terus meningkat sebesar 32,7% untuk setiap tahunnya. (Wicaksana & Nasvian,

OKTAL : Jurnal Ilmu Komputer dan Science
Volume 3, No. 4, April 2024
ISSN 2828-2442 (media online)
Hal 865-874

2022). Meski begitu, pada awalnya industri *e-sports* dipandang sebelah mata dan dianggap sebagai kegiatan yang tidak bermanfaat dan hanya membuang waktu saja oleh kalangan para dewasa khusunya orang tua. Tidak menutup kemungkinan bahwa adanya *e-sports* memberikan berbagai macam dampak baik positif maupun negatif. Beragam tanggapan mengenai industri *e-sports* tersebar di banyak media sosial, salah satunya *twitter*.

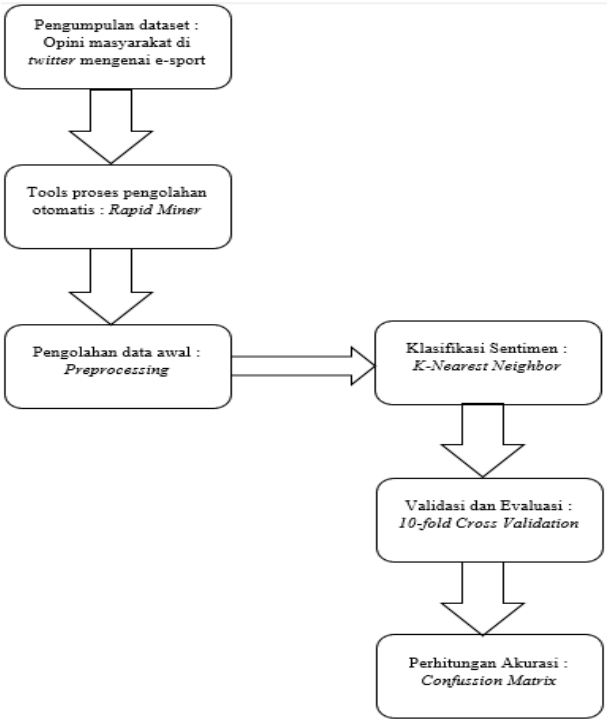
Twitter merupakan media sosial besar di Indonesia yang banyak digunakan masyarakat untuk berbagai keperluan seperti curhat, beropini, atau mengobrol bersama teman. (Hidayat et al., 2021) Opini di *twitter* bisa dianggap informasi namun masih kurang terstruktur, sehingga masyarakat kesulitan memahami dan menentukan antara informasi yang benar dan salah. Opini tersebut dapat diolah menjadi suatu informasi terstruktur dengan melakukan analisis sentimen.

Analisis sentimen ialah teknik mengekstrak data teks guna mendapatkan informasi tentang sentimen bernilai positif, netral ataupun negatif. Analisis sentimen dilakukan untuk melihat opini terhadap suatu masalah atau digunakan untuk mengidentifikasi kecenderungan dari suatu permasalahan. (Simorangkir & Lhaksmana, 2018). Pada penelitian ini, bertujuan untuk mengetahui kecenderungan masyarakat mengenai industri *e-sports*, yang diharapkan hasil nya bisa mengatasi kekeliruan informasi tentang perkembangan industri *e-sports* dan dapat merubah stigma masyarakat serta mengurangi perdebatan. Dalam melakukan analisis sentimen diperlukan sebuah metode yang mampu mengklasifikasikan sentimen.

Solusi metode klasifikasi yang digunakan pada penelitian ini yaitu *K-Nearest Neighbor*, yang merupakan algoritma untuk mengklasifikasi objek baru berdasarkan atribut dan data *training*. Tingkat akurasi sentimen diukur menggunakan tabel *confussion matrix*. *Confussion Matriks* merupakan metode *external* evaluasi yang berisi informasi aktual dan dapat diprediksi, dimana kinerja sistem dapat di evaluasi menggunakan data dalam matriks. (Kusumawardhana, 2019) Berdasarkan uraian diatas, penulis tertarik untuk melakukan penelitian sebagai pemenuhan mata kuliah skripsi berjudul "ANALISIS SENTIMEN TERHADAP INDUSTRI *E-SPORTS* PADA MEDIA SOSIAL *TWITTER* DENGAN MENGGUNAKAN METODE *K-NEAREST NEIGHBOR*".

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian



Gambar 1. Tahapan Penelitian

OKTAL : Jurnal Ilmu Komputer dan Science
Volume 3, No. 4, April 2024
ISSN 2828-2442 (media online)
Hal 865-874

Dari gambar diatas di jelaskan pada tahapan pertama yaitu pengumpulan dataset, dengan cara melakukan *crawling twitter* menggunakan *rapid miner*. Data yang diambil sebanyak 200 komentar mengenai *e-sports*, yang kemudian komentar tersebut diintegrasikan menjadi dataset. Tahapan selanjutnya yaitu *tools* proses pengolahan otomatis yang akan menggunakan *rapid miner*.

Tahap ketiga yaitu pengolahan data awal atau teks *preprocessing* yang mencakup *cleaning* yaitu proses menghilangkan atribut atau karakter tidak perlu seperti tanda baca, simbol dan lain sebagainya, *case folding* yaitu proses mengubah data teks menjadi *lowercase*, normalisasi yaitu proses memperbaiki kata menjadi kata baku yang mengandung arti, *tokenization* yaitu proses memotong kata menjadi tunggal atau memecah per kata, *stopword removal* yaitu proses membuang kata tidak penting yang tidak mengandung sentimen seperti kata hubung, dan terakhir *stemming* yaitu proses membuang imbuhan menjadi kata dasar.

Tahap keempat yaitu klasifikasi sentimen dengan metode yang digunakan yaitu *K-Nearest Neighbor* yang terbukti mencapai hasil akurasi yang baik dan sesuai dengan perhitungan yang diterapkan dalam sebuah aplikasi. Untuk melakukan perhitungan kedekatan antara dua kasus akan menggunakan rumus *consine similarity.* Tahap kelima yaitu validasi dan evaluasi dengan menerapkan *10-fold cross validation*. Kemudian tahapan terakhir yaitu menghitung akurasi dengan memanfaatkan rumus *confussion matrix.*

2.2 Metode Pengumpulan Data

Penelitian ini menggunakan teknologi data *mining* dalam melakukan klasifikasi analisis sentimen dengan sumber data berasal dari komentar *twitter* untuk menentukan apakah bersifat positif atau negatif. Penulis menggunakan metode *K-Nearest Neighbor* untuk melakukan analisis sentimen terhadap industri *e-sports* pada media sosial *twitter*. Data yang akan digunakan yaitu data *training* yang diambil dari 150 komentar *twitter* minggu sebelumnya sebagai acuan dan data *testing* yang juga akan diambil dari komentar *twitter* secara *realtime* pada saat proses berlangsung. *Crawling* data akan dilakukan secara otomatis dengan *tools rapid miner* dengan cara menghubungkan *tools* pada *Application Inferface (API) twitter.* Data yang diperoleh dari hasil *crawling* di *twitter* akan di proses terlebih dahulu pada tahapan-tahapan dalam *preprocessing* sampai kemudian dapat di klasifikasikan. Kata kunci yang digunakan ada 3 yaitu *esport*, *e-sport*, dan *esports.*

2.3 Metode Analisis

Analisis sentimen ialah pengolahan bahasa natural yang berperan menganalisis pendapat dan juga dapat diartikan sebagai suatu pandangan, keadaan serta perilaku seseorang dalam tulisan maupun lisan mereka. Metode analisis yang penulis terapkan yaitu metode analisis sentimen yang mana metode ini merupakan metode analisis data kualititatif, pada metode ini data yang dikumpulkan adalah data sekunder yang bersumber dari media sosial *twitter*.

Proses pengumpulan data dibatasi hanya mengenai topik ataupun pendapat masyarakat mengenai *e-sports* pada media sosial *twitter* menggunakan metode *crawling twitter* di *rapid miner*. Data yang sudah ada nantinya akan menjadi acuan untuk menguji klasifikasi sentimen positif dan negatif.

2.4 Analisis Sentimen

Analisis sentimen atau *opinion mining* merupakan proses memahami, mengekstrak dan mengolah data tekstual untuk mendapat informasi sentimen yang terkandung dalam suatu opini. (Simorangkir & Lhaksmana, 2018). Tujuan dari analisis sentimen adalah untuk mendefinisikan alat otomatis yang mampu mengekstrak informasi subjektif dari teks dalam bahasa alami, seperti pendapat dan sentimen, sehingga menciptakan pengetahuan yang terstruktur dan dapat ditindaklanjuti untuk digunakan baik oleh *decision support system* atau *decision maker* (Jabal Tursina, 2019)

Dalam penelitian ini penulis menganalisis sentimen masyarakat mengenai industri *e-sports*. Dataset yang digunakan bersumber dari komentar masyarakat di *twitter* dengan data yang dibatasi hanyalah mengambil komentar berbahasa Indonesia dalam bentuk teks tanpa mengandung gambar. Proses pengumpulan data dilakukan otomatis dengan metode *crawling* yang kemudian disimpan di *microsoft excel*. Pada gambar 2 diperlihatkan hasil *crawling* yang menjadi dataset untuk penelitian.

OKTAL : Jurnal Ilmu Komputer dan Science
Volume 3, No. 4, April 2024
ISSN 2828-2442 (media online)
Hal 865-874

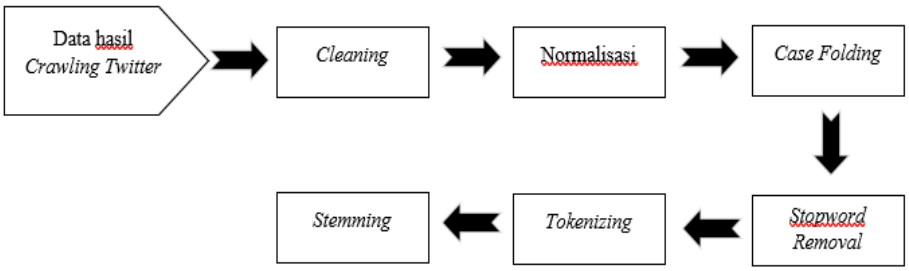
	A	B	
1	From-User	Text	Id
2	gio	@ChelseaFC Bubar aja bang jgn jadi klub bola, jadi tim e-sport aja mendingan	1626973255520 35
3	ilona	13. suka main game (pernah masuk esport)	1629713009789 78
4	Godzilla Goreng	@Txtmlbb gak semua atlit esport itu cringe kek gini ya geess	1629707002065 85
5	BerbagiViral09	Perkuat Kabar Kolaborasi Bigetron Esport dan Persib Bandung, Babyla dan Mute Tiba di Markas Pangeran Biruhttps://t.co/bFDDr5fiNf	1629682596170 04
6	SAHABAT GANJAR OF	Inilah pemenang Turnamen Mobile Legends SAGA Esports di Kabupaten Sumedang, Jawa Barat.Juara 1 Main Hati EsportJuara 2 Edge LegendJuara 3 Unsap EsportSelamat kepada para pemenang	1629666612986 15
7	sera – phew! 。	@mlbbfess Dukung esport doang turnamen ga diliat ilmunya xixi	1629663295597 84
8	C	GATAU ANJIR sisi gelap dunia esport https://t.co/MqRwns2Aqg	1629651490402 09
9	Fobis Media	Komposisi ONIC Esports 2023, Salah Satu tim Mobile Legends Terbaik Indonesia https://t.co/JgCqpYPu1T #gadget #otomotif #game	1629651103221 69
10	immort4l	RT @exbadut: @Txtmlbb Emg 2 lembaga itu bener menyuarakan rakyt toh? Esport biar jdi milik komunitas aja, ga usah banyak bacot mau nyari la...	1629651075522 50
11	tvOnenews	Tak Dapat Untung, Persib Bandung Tutup Rapat Pintu untuk Esport https://t.co/xQavekjVXs	1629589187400 97
12	Bigmom	Ngeri bgt ketemu duo dari esport AGS gendong2 bgt gilak btw ags esport tuh dari daerah mana y https://t.co/kdkGXjbI5Y	1629555132940 82
13	iRen	@mlbbfess semoga rans esport masuk final	1629529147008 76
14	Miuu	Jago bgt ak cocok msk tim esport pinggir jurang	1629525316904 58
15	adt	paling nungguin bgt klo jaya esport mabarr https://t.co/xDmtYvdHKB	1629521939537 76
16	Lei	Kalo ngesimp player Esport normal ga, soalnya gua mau bar bar nih	1629521739385 55
17	ICEAMERICANO	Yang awalnya cuma iseng" main game dan sampe jadi manager team esport, walaupun cuma team lokal dan ga kebayang sampe jadi side job dan sedih banget kalo harus berhenti. People come	1629519799490 60
18	AlurInformasi	Evos Legends Mengalami Kekalahan Pertama Menghadapi Bigetron Alpha#evos #btr #Bigetron#mpl #Olahraga #esporthttps://t.co/Kmh1XEuYbs https://t.co/BOnPce4pRP	1629507383738 39
19	hunter | freetag	persija esport ngerii	1629470349808 25
20	panggil saja aditya	@memefess Sepakbola dipegang pemerintah hancur bang. Banyak mafia, wasit ngawur, dan masih banyak lagi. Esport udah bagus, enggak ada masalah apa-apa.	1629437439101 27
21	djamesss	@GamesMenfess Sok ngancem2 , Baru dateng aja di esport indo ngantur.	1629406679581 49
22	bae	Alangkah baiknya urusan geming2 dan Esport ini diurus yg muda2 sj drpd bumer2 gaptek kek kmtl https://t.co/EYE77x7E3y	1629395160315 44
23	cfp	Gila ya gaji pemain2 esport ini.. sampi bisa beli BMW M4..	1629385034980 55

Activate Windows

Gambar 2. Dataset Penelitian

2.5 *Text Preprocessing*

Preprocessing merupakan tahapan awal dan tahap paling penting dari proses *text mining* untuk mengubah data sesuai dengan format yang akan dibutuhkan pada proses selanjutnya. (Kusumawardhana, 2019). *Preprocessing* adalah proses untuk mempersiapkan data mentah sebelum dilakukan proses lain. Untuk itu, *preprocessing* ini berfungsi sebagai proses pembersihan data atau untuk mengatasi *noise* dan mengeliminasi data yang tidak sesuai menjadi bentuk yang lebih mudah di proses oleh sistem (Safitri, 2020).



Gambar 3. Alur Proses *Preprocessing*

Penjelasan tahapan *preprocessing* diatas antara lain :

a. *Cleaning* : untuk mengeliminasi karakter atau atribut tidak diperlukan pada dataset seperti tanda baca, simbol, angka, *emoticon*, *mention*, *hashtag*, *link*, *url*, *retweet* dan *noise* lainnya. Contohnya (.,?!'::"-_=+//*{}[]><^&%$~)
b. Normalisasi : yaitu untuk memperbaiki kata-kata menjadi kata baku yang sesuai dengan Kamus Besar Bahasa Indonesia juga membenarkan kata singkatan. Misal, 'dgn' menjadi 'dengan'.
c. *Case Folding* : proses untuk mengubah setiap kata pada dokumen menjadi format yang sama, yaitu *lower case* atau huruf kecil.
d. *Stopword Removal* : yaitu proses membuang kata pada dokumen yang tidak mengandung makna atau tidak begitu penting. Kata-kata yang dibuang biasanya kata hubung atau yang termasuk *stopword*, yaitu kata yang diabaikan, yang mempunyai fungsi namun tidak memiliki arti. Misal, 'karena', 'atau', 'bagaimana'.
e. *Tokenizing* : dilakukan untuk memotong data dokumen menjadi bagian-bagian kecil *token* atau memecahkan kalimat menjadi per kata agar mendapat kata yang memiliki nilai.
f. *Stemming* : proses penghapusan kata imbuhan yang terdapat pada dataset guna mendapat kata dasar yang mudah diolah oleh sistem.

OKTAL : Jurnal Ilmu Komputer dan Science
Volume 3, No. 4, April 2024
ISSN 2828-2442 (media online)
Hal 865-874

2.6 Metode *K-Nearest Neighbor*

K-Nearest Neighbor adalah algoritma untuk mengklasifikasi objek baru berdasarkan atribut dan data latih, dimana hasil dari sampel uji yang baru diklasifikasikan berdasarkan mayoritas dari kategori pada *K-Nearest Neighbor*. Algoritma *K-Nearest Neighbor* menggunakan klasifikasi ketetanggaan sebagai nilai prediksi dari sampel uji yang baru. (Rizki, 2019) Dijelaskan bahwa algoritma *K-Nearest Neighbor* adalah pendekatan untuk mencari kasus baru dengan kasus lama, yaitu berdasarkan pada kecocokan bobot dari sejumlah fitur yang ada. (Tanggu Mara et al., 2021)

Adapun algoritma *K-Nearest Neighbor* dapat dijelaskan dengan keterangan sebagai berikut :

a. Menentukan parameter nilai k jumlah tetangga terdekat.
b. Menghitung bobot setiap *term* dengan *TF-IDF*.
c. Menghitung kemiripan antar dokumen dengan persamaan *cosine similarity*. Berikut rumus persamaan nya :

$$CosSim(q, dj) = \frac{dj.q}{|dj|.|q|} = \frac{\sum_{i=1}^{t}(wij.wiq)}{\sqrt{\sum_{i=1}^{t} w_{ij}^2} \cdot \sqrt{\sum_{i=1}^{t} w_{iq}^2}}$$

CosSim (q,dj) => Nilai kemiripan antara dokumen (q) dengan dokumen latih ke j (d1)
T => Jumlah term (kata)
D => dokumen
Q => kata kunci (query)
Wij => Bobot term (kata) ke i pada dok. latih j
Wiq => Bobot term (kata) ke i pada dok. uji q

d. Urutkan hasil perhitungan *cosine similarity* dari besar ke kecil.
e. Ambil sebanyak k yang paling tinggi kemiripannya dengan dokumen yang diklasifikasikan, tentukan kelasnya.
f. Cari jumlah terbanyak dari tetanggaan terdekat. Kemudian tetapkan kategori.

2.7 *Confussion Matrix*

Dalam mengukur tingkat akurasi sistem orientasi sentimen, maka digunakan tabel *confussion matrix*. *Confussion Matrix* adalah sebuah metode yang biasa digunakan untuk perhitungan akurasi, *recall*, *precision*, dan *error rate*. (Herdhianto, 2020). Menurut (Kusumawardhana, 2019), *Confussion Matriks* merupakan metode *external* evaluasi yang berisi informasi aktual dan dapat diprediksi, dimana kinerja sistem dapat di evaluasi menggunakan data dalam matriks.

Berikut menunjukan tabel dari *confussion matrix*.

Tabel 1. *Confussion Matrix*

Kelas Sebenarnya	**Kelas Hasil Prediksi**	
	Positif = 1	**Negatif = 0**
Positif = 1	TP	FN
Negatif = 0	FP	TN

Keterangan untuk tabel diatas dinyatakan sebagai berikut:

a. *True Positive* (TP), yaitu jumlah dokumen dari kelas 1 yang benar dan diklasifikasikan sebagai kelas 1.
b. *True Negative* (TN), yaitu jumlah dokumen dari kelas 0 yang benar diklasifikasikan sebagai kelas 0.
c. *False Positive* (FP), yaitu jumlah dokumen dari kelas 0 yang salah diklasifikasikan sebagai kelas 1.
d. *False Negative* (FN), yaitu jumlah dokumen dari kelas 1 yang salah diklasifikasikan sebagai kelas 0.

Perhitungan akurasi dirumuskan sebagai berikut:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

OKTAL : Jurnal Ilmu Komputer dan Science
Volume 3, No. 4, April 2024
ISSN 2828-2442 (media online)
Hal 865-874

3. ANALISA DAN PEMBAHASAN

Sebagian besar proses pada penelitian ini dilakukan secara otomatis dengan *rapid miner*. *Rapid miner* merupakan solusi untuk melakukan analisis dengan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik. Penggunaan *rapid miner* untuk mempermudah dan mempersingkat proses analisis sentimen, meski begitu ada beberapa tahapan yang dilakukan secara manual. Adapun tujuan penelitian ini adalah untuk melakukan penerapan algoritma *K-Nearest Neighbor (K-NN)* dan memperoleh hasil akurasi terbaik dalam mengklasifikasikan sentimen yang hasilnya dapat menjadi sebuah informasi. (Masturoh, 2021).

3.1 *Preprocessing* Data

Alur tahapan sistem pada *preprocessing* data secara berurut meliputi *cleaning*, normalisasi, *case folding*, *stopword removal*, *tokenizing* dan *stemming*. Berikut sampel hasil *preprocessing* yang tertera pada tabel 2.

Tabel 2. Hasil *Preprocessing*

Tahapan	Hasil
Data Tweet	@ChelseaFC Bubar aja bang jgn jadi klub bola, jadi tim e-sport aja mendingan
Cleaning	Bubar aja bang jgn jadi klub bola jadi tim esport aja mendingan
Normalisasi	Bubar saja bang jangan jadi klub bola jadi tim esport saja mendingan
Case Folding	bubar saja bang jangan jadi klub bola jadi tim esport saja mendingan
Stopword removal	bubar bang klub bola tim esport mendingan
Tokenizing	bubar || bang || klub || bola || tim || esport || mendingan
Stemming	bubar || bang || klub || bola || tim || esport || mending

3.2 *Labelling* Kelas Sentimen

Pelabelan sentimen dilakukan secara manual dengan menilai data dalam kelompok sentimen berdasarkan hasil skor. Untuk mengetahui kategori suatu kata disesuaikan dengan kamus *lexicon* yang berisi kumpulan-kumpulan kata positif dan negatif. Contoh hasil perhitungan kelas sentimen terdapat pada tabel dibawah.

Tabel 3. Sampel Perhitungan Kelas Sentimen

Tweet	Kata Positif	Kata Negatif	Skor	Sentimen
bubar bang klub bola tim esport mending	-	bubar	-1	Negatif
suka main game masuk esport	suka main	-	2	Positif

Dari proses *labelling* kelas sentimen data *training* didapat hasil sentimen positif sebanyak 82 data dan sentimen negatif sebanyak 68 data.

Tabel 4. Hasil Pelabelan Kelas Sentimen

Kelas sentimen	Jumlah
Positif	82
Negatif	68
Total data	**150**

3.3 Pembobotan *Term TF-IDF*

Setelah melakukan *labelling* pada dataset, proses selanjutnya yaitu tahap pembobotan *TF-IDF*. Tujuannya untuk memberikan bobot setiap kata (*term*) pada setiap data pada dokumen teks. Pembobotan *term TF-IDF* akan dilakukan secara otomatis. Contoh hasil perhitungan *TF-IDF* dapat dilihat pada tabel 5.

OKTAL : Jurnal Ilmu Komputer dan Science
Volume 3, No. 4, April 2024
ISSN 2828-2442 (media online)
Hal 865-874

Tabel 5. Perhitungan TF-IDF

TERM	TF (d1)	TF (d2)	TF (d3)	DF	N/DF	IDF (log N/DF)
bubar	1	0	0	1	3	0.477121255
bang	1	0	0	1	3	0.477121255
klub	1	0	0	1	3	0.477121255
bola	1	0	0	1	3	0.477121255
tim	1	0	0	1	3	0.477121255
esport	1	1	1	3	1	0
mending	1	0	0	1	3	0.477121255

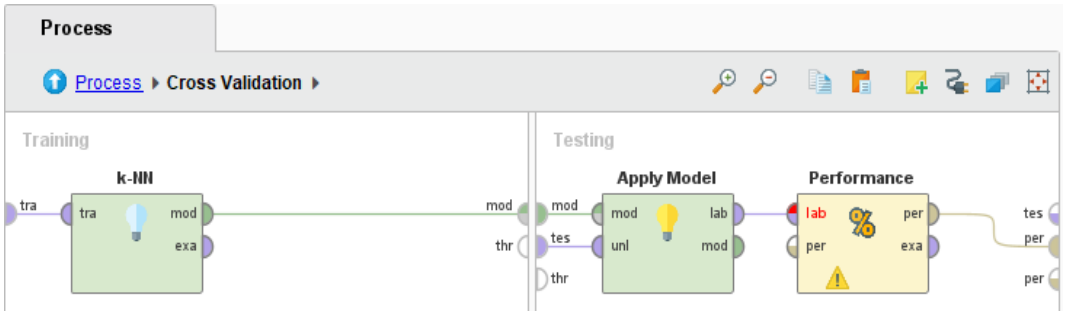
Setelah melalui proses dalam menemukan nilai *TF-IDF*, maka kita bisa menentukan bobot suatu kata dengan mengkalikan nilai *term frequency* dan *inverse document frequency*. Contoh perhitungan bobot diuraikan pada tabel di bawah.

Tabel 6. Perhitungan Bobot (*Weight*)

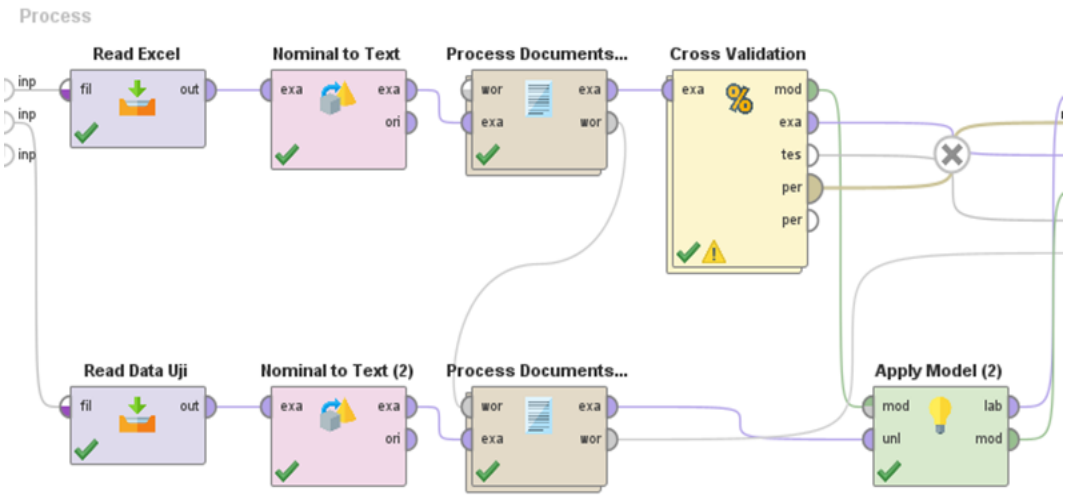
TF.IDF		
d1	**d2**	**d3**
0.477121255	0	0
0.477121255	0	0
0.477121255	0	0

3.4 Klasifikasi *K-Nearest Neighbor*

Proses klasifikasi dan perhitungan *cosine similarity* dilakukan secara otomatis dengan *rapid miner*. Dalam penelitian ini penulis akan menguji nilai k=1 sampai dengan k=10. Nilai akurasi tertinggi nantinya akan dipilih sebagai model pengklasifikasian sentimen pada *data testing.*



Gambar 4. Proses *K-NN* pada *Rapid Miner*



Gambar 5. Keseluruhan Tahapan Analisis Sentimen

OKTAL : Jurnal Ilmu Komputer dan Science
Volume 3, No. 4, April 2024
ISSN 2828-2442 (media online)
Hal 865-874

4. IMPLEMENTASI

4.1 Uji Validasi dan Perhitungan Akurasi

Berikut adalah hasil uji validasi *10-Fold Cross Validation* dari masing-masing nilai k yang di hitung secara otomatis.

Tabel 7. Tabel Uji Validasi

Nilai K -	*Accuracy*	*Precision*	*Recall*	*AUC*
1	59.33 %	61.81 %	69.86 %	0.500
2	59.33 %	61.81 %	69.86 %	0.580
3	64.67 %	66.21 %	73.33 %	0.672
4	66.00 %	68.08 %	74.58 %	0.662
5	64.00 %	64.86 %	76.67 %	0.695
6	**66.00 %**	**67.34 %**	**78.06 %**	**0.726**
7	64.00 %	64.26 %	80.42 %	0.740
8	64.00 %	65.32 %	76.67 %	0.724
9	62.00 %	65.65 %	71.67 %	0.710
10	65.33 %	68.59 %	74.44 %	0.721

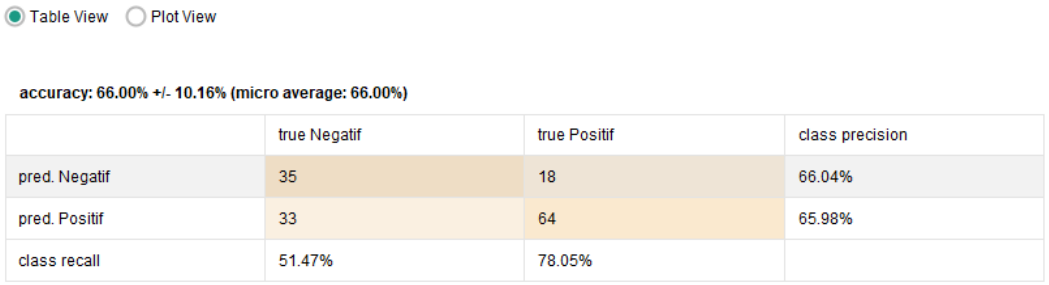
Berdasarkan pengujian nilai k pada tabel diatas terdapat beberapa akurasi bernilai sama, namun pada nilai k=6 memiliki nilai AUC lebih tinggi sehingga nilai k=6 mendapat hasil tertinggi dengan akurasi sebesar 66.00 %, presisi dengan nilai 67.34 %, *recall* 78.06 % dan AUC dengan nilai 0.762. Dengan begitu, nilai k=6 dapat dipilih sebagai model pengklasifikasian sentimen pada data *testing*.

Berikut perhitungan akurasi menggunakan rumus *confussion matrix*.

$$Accuracy = \frac{True\ Positive + True\ Negative}{Total\ Data\ Matrix} \text{ X } 100\ \%$$

$$Accuracy = \frac{64+35}{64+18+33+35} \text{ X } 100\ \%$$

$$Accuracy = \frac{99}{150} \text{ X } 100\ \% = 66\%$$



Gambar 6. Hasil Akurasi Nilai K=6

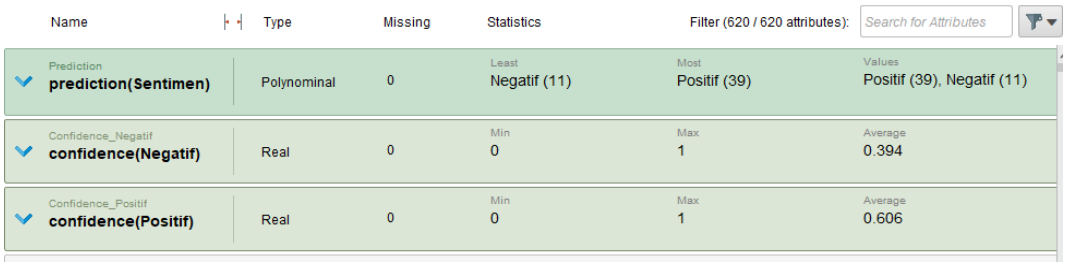
4.2 Hasil Analisis Sentimen

Data *testing* yang dipakai berjumlah 50 dataset yang bersumber dari *twitter*. Setelah melewati tahap preprocessing untuk mendapat data yang terstruktur, maka dilakukan pelabelan sentimen dengan memanfaatkan model klasifikasi yang telah dibangun pada data *training*. Dari proses tersebut, didapat hasil pelabelan sentimen pada data *testing* sebagai berikut:

Tabel 8. Hasil Sentimen Data Testing

Kelas Sentimen	Jumlah
Positif	39
Negatif	11
Total	**50**

OKTAL : Jurnal Ilmu Komputer dan Science
Volume 3, No. 4, April 2024
ISSN 2828-2442 (media online)
Hal 865-874



Name	Type	Missing	Statistics	Filter (620 / 620 attributes): Search for Attributes	
Prediction **prediction(Sentimen)**	Polynominal	0	Least Negatif (11)	Most Positif (39)	Values Positif (39), Negatif (11)
Confidence_Negatif **confidence(Negatif)**	Real	0	Min 0	Max 1	Average 0.394
Confidence_Positif **confidence(Positif)**	Real	0	Min 0	Max 1	Average 0.606

Gambar 7. Hasil Statistik Data Testing

Row No.	prediction(S...	confidence(...	confidence(...
1	Positif	0.500	0.500
2	Negatif	0.670	0.330
3	Negatif	1	0
4	Positif	0.337	0.663
5	Positif	0.335	0.665
6	Negatif	0.512	0.488
7	Positif	0.165	0.835
8	Negatif	0.505	0.495
9	Positif	0.339	0.661
10	Positif	0.497	0.503
11	Negatif	0.666	0.334
12	Positif	0.336	0.664
13	Positif	0.331	0.669

ExampleSet (50 examples, 3 special attributes, 617 regular attributes)

Gambar 8. Hasil Klasifikasi Data Testing

5. KESIMPULAN

Analisis sentimen terhadap industri *e-sports* menggunakan 200 dataset yang terbagi ke dalam 150 data training dan 50 data testing yang semuanya bersumber dari media sosial *twitter*. Dari pengujian pada data *training* yang selanjutnya dijadikan model pada pengklasifikasian data *testing*, didapat hasil bahwa kelas sentimen positif mendapat keunggulan sebanyak 39 opini dibandingkan kelas negatif yang mendapat kelas sentimen berjumlah 11 opini. Dengan kata lain masyarakat Indonesia saat ini dalam menanggapi topik permasalahan mengenai *e-sports* cenderung lebih memberikan stigma positif dan respon yang baik. Dari penelitian ini juga menunjukkan kinerja metode *K-Nearest Neighbor* yang dapat melakukan pengujian dengan baik dengan memanfaatkan uji validasi *10-fold cross validation* dan *cosine similarity* sebagai perhitungan kemiripan dokumen. Dari perbandingan nilai k=1 sampai dengan k=10 yang diuji, didapat hasil nilai k=6 sebagai hasil akurasi tertinggi yang dijadikan model pengklasifikasian terhadap data *testing*. Adapun hasil akurasi menunjukkan nilai sebesar 66.00%, presisi dengan nilai 67,34%, recall mendapat nilai 78,06% dan nilai AUC sebesar 0,762. Pada perhitungan akurasi menggunakan *confussion matrix* pada nilai k=6 juga mendapat hasil yang sesuai dengan perhitungan otomatis, yaitu 66%.

OKTAL : Jurnal Ilmu Komputer dan Science
Volume 3, No. 4, April 2024
ISSN 2828-2442 (media online)
Hal 865-874

REFERENCES

Herdhianto, A. (2020). *Sentiment Analysis Menggunakan Naïve Bayes Classifier (NBC) Pada Tweet Tentang Zakat*. http://repository.uinjkt.ac.id/dspace/handle/123456789/53661

Hidayat, T. F. T., Garno, G., & Ridha, A. A. (2021). Analisis Sentimen Opini Pemindahan Ibu Kota Pada Twitter Dengan Metode Support Vector Machine. *Jurnal Ilmu Komputer*, *14*(1), 49. https://doi.org/10.24843/jik.2021.v14.i01.p06

Jabal Tursina, M. (2019). Sentimen Analisis Sistem Zonasi Sekolah Pada Media Sosial Youtube Menggunakan Metode K-Nearest Neighbor Dengan Algoritma Levenshtein Distance. *Universitas Islam Negeri Syarif Hidayatullah Jakarta*.

Kusumawardhana, A. (2019). *KLASIFIKASI SENTIMEN TERHADAP TOKOH PUBLIK PADA TWITTER MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR*.

Masturoh, S. (2021). *Analisis Sentimen Terhadap E-Wallet Dana Pada Ulasan Google Play Menggunakan Algoritma K-Nearest Neighbor*. https://doi.org/10.33480/pilar.v17i1.2182

Rizki, M. M. (2019). Analisis sentimen terhadap produk otomotif dari twitter menggunakan kombinasi algoritma k-nearest neighbor dan pendekatan lexicon (studi kasus: mobil toyota). *Repository.Uinjkt.Ac.Id*. http://repository.uinjkt.ac.id/dspace/handle/123456789/48643

Safitri, R. N. (2020). *Analisis Sentimen Review Pelanggan Hotel Menggunakan Metode K-Nearest Neighbor (K-NN) (Studi Kasus : Hotels.com, Booking.com, Agoda.com)*.

Simorangkir, H., & Lhaksmana, K. M. (2018). Analisis Sentimen pada Twitter untuk Games Online Mobile Legends dan Arena of Valor dengan Metode Naïve Bayes Classifier. *E-Proceeding of Englineering*, *5*(3), 8131–8140. https://openlibrary.telkomuniversity.ac.id/pustaka/files/144621/jurnal_eproc/analisis-sentimen-pada-twitter-untuk-games-online-mobile-legends-dan-arena-of-valor-dengan-metode-na-ve-bayes-classifier.pdf

Tanggu Mara, A., Sediyono, E., & Purnomo, H. (2021). Penerapan Algoritma K-Nearest Neighbors Pada Analisis Sentimen Metode Pembelajaran Dalam Jaringan (DARING) Di Universitas Kristen Wira Wacana Sumba. *Jointer - Journal of Informatics Engineering*, *2*(01), 24–31. https://doi.org/10.53682/jointer.v2i01.30

Wicaksana, F. A., & Nasvian, M. F. (2022). KOMUNIKASI, KOORDINASI, DAN KERJASAMA DALAM GAME KOMPETITIF MOBILE LEGEND. *Jurnal Ilmiah Indonesia*, *7*(8.5).