



Penerapan Algoritma K-Nearest Neighbor (K-NN) Untuk Analisis Sentimen Publik Terhadap Pembelajaran Daring

Jepi Supriyanto^{1*}, Debby Alita², Auliya Rahman Isnain³

^{1*,2,3}Informatika, Universitas Teknokrat Indonesia, Indonesia

^{1*}japi.supriyanto@gmail.com, ²debbyalita@teknokrat.ac.id, ³auliyarahman@teknokrat.ac.id

Submitted : 23 February 2023 | Accepted : 28 February 2023 | Published : 15 March 2023

Abstrak: Penelitian ini dilakukan untuk menerapkan algoritma KNN (K-Nearest Neighbor) dalam melakukan sentimen analisis pengguna *Twitter* tentang isu terkait kebijakan pemerintah tentang Pembelajaran Daring. penelitian menggunakan data *Tweet* sebanyak 1825 data tweet Bahasa Indonesia data dikumpulkan sejak tanggal 1 Februari 2020 sampai dengan 30 September 2020. Menggunakan library python yaitu *Tweepy*. pembobotan kata menggunakan TF-IDF, akan dilakukan klasifikasi nilai sentimen ke dalam dua kelas yaitu positif dan negatif. Setelah dilakukan pengujian dengan K sebanyak 20 didapatkan hasil akurasi tertinggi terdapat Pada saat K = 10 dengan nilai akurasi 84,65% dengan presisi mencapai 87%, *recall* 86% *f measure* 87% serta error rate mencapai 0,12% dan di dapatkan pula kecenderungan opini publik terhadap Pembelajaran Daring Cenderung Positif

Kata Kunci: *Confusion Matrix; K Nearest Neighbor; Pembelajaran Daring; Sentimen Analisis; TF-IDF;*

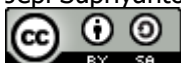
Abstract: This research was conducted to apply the KNN (K-Nearest Neighbor) algorithm in conducting sentiment analysis of Twitter users on issues related to government policies regarding Online Learning. Research using Tweet data as much as 1825 Indonesian tweet data data were collected from February 1, 2020 to September 30, 2020. Using the python library, Tweepy. word weighting using TF-IDF, will be classified into two classes of sentiment values, positive and negative. After testing with K of 20, the highest accuracy results were obtained when K = 10 with an accuracy value of 84.65% with a precision of 87%, a recall of 86% *f measure* 87% and an error rate of 0.12% and a tendency was also obtained. public opinion on online learning tends to be positive

Keywords: *Confusion Matrix; K Nearest Neighbor; Online Learning; Sentiment Analysis; TF-IDF;*

1. PENDAHULUAN

Sejak diumumkan kasus positif *Corona virus* (Covid-19) di Indonesia pada tanggal 2 maret 2020, sampai dengan tanggal 31 juli 2020 tercatat total 108.376 kasus dikonfirmasi dengan rincian pasien dalam perawatan mencapai 37,338 orang yang tersebar di seluruh rumah sakit yang ada di indonesia, 65,907 pasien yang dinyatakan sembuh dan tercatat kasus meninggal dunia mencapai angka 5,131 orang. Pandemi global yang terjadi di Indonesia membuat banyak pihak berupaya ikut berperan serta dalam mengatasi. Para dokter umum dan spesialis angkat bicara bersama guna memberi penjelasan singkat kepada masyarakat maupun himbauan agar menjaga kebersihan diri dan lingkungan sekaligus tak banyak keluar rumah[1]. Salah satu dampak pandemi *Corona virus* ialah terhadap pendidikan di seluruh dunia, yang mengarah kepada penutupan sekolah, madrasah, universitas, dan pondok pesantren, Sehubungan dengan perkembangan tersebut, Kementerian Pendidikan dan Kebudayaan (Kemendikbud) turut mengambil kebijakan sebagai panduan dalam menghadapi penyakit tersebut di

Jepi Supriyanto: *Penulis Korespondensi



Copyright © 2023, Jepi Supriyanto, Debby Alita, Auliya Rahman Isnain.



tingkat satuan pendidikan dengan mengadakan pembelajaran jarak jauh menanggapi kebijakan Kementerian Pendidikan dan Kebudayaan, banyak menuai opini-opini dikalangan masyarakat yang mereka tuangkan melalui media sosial terutama di Twitter Untuk menafsirkan dan memahami opini tersebut, diperlukan algoritma dan program untuk mengolah data informasi dan opini, juga menganalisa opini.

Sentiment Analisis atau biasa disebut *opinion mining* merupakan salah satu cabang penelitian *Text Mining*. *Opinion mining* adalah riset komputasional dari opini, *sentimen* dan emosi yang diekspresikan secara tekstual. Analisis sentimen atau penambangan opini adalah suatu bidang studi untuk menganalisis pendapat orang terhadap entitas seperti produk, layanan, organisasi, individu, masalah, peristiwa, dan topik. Analisis sentimen ini berfokus pada pendapat seseorang yang mengekspresikan atau menyiratkan sentimen Positif atau negatif, kebanyakan analisis sentimen ini berkaitan dengan orang-orang di media sosial Selain polaritas Positif dan negatif, terkadang polaritas juga dianggap sebagai kisaran dimana suatu dokumen dapat berisi pernyataan yang memiliki polaritas campuran Mejova, (2009).

TFIDF adalah teknik pengambilan informasi yang membebani *term frequency* dan dokumen inversnya (IDF). Setiap kata atau istilah memiliki skor TF dan IDF bobot dari istilah disebut sebagai TFIDF. Algoritma TFIDF digunakan untuk menimbang *keyword* dalam setiap dokumen dan menghitung seberapa banyak kemunculannya dalam setiap dokumen. K-Nearest Neighbor (KNN) adalah salah satu algoritma klasifikasi supervised learning yang digunakan untuk mengklasifikasikan objek berdasarkan atribut kelas dan data training[3]. Konsep dasar dari K-NN adalah mencari jarak terdekat antara data yang akan dievaluasi dengan K tetangga terdekatnya dalam data pelatihan. Penghitungan jarak dilakukan dengan konsep Euclidean. Jumlah kelas yang paling banyak dengan jarak terdekat tersebut akan menjadi kelas dimana data evaluasi tersebut berada Tahapan awal pada KNN ini adalah menentukan nilai K, misal k=3 artinya 3 dokumen terdekat dengan dokumen uji yang akan diambil.

Berdasarkan referensi tersebut Penulis pada penelitian ini akan melakukan *analisis sentimen* para pengguna *Twitter* terhadap pembelajaran Daring. Dengan *input* berupa data *tweet* dalam Bahasa Indonesia, akan dilakukan klasifikasi dengan algoritma KNN (*K-Nearest Neighbor*) untuk menentukan apakah *tweet* tersebut bersentimen Positif, atau negatif.

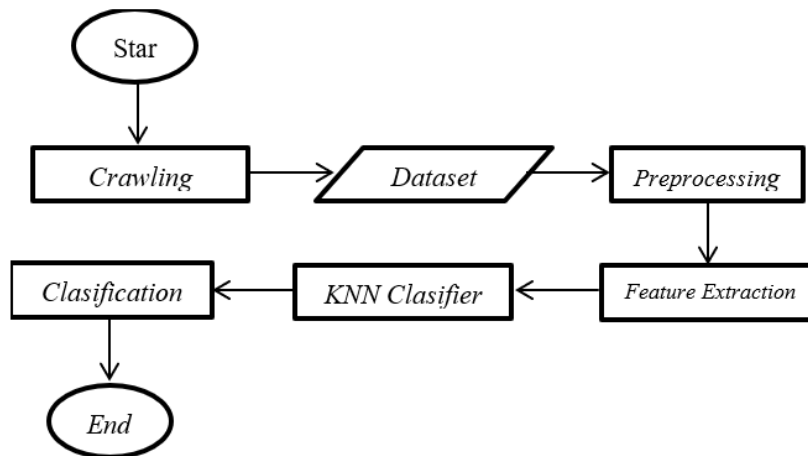
2. METODE PENELITIAN

Dalam melakukan *text mining*, *text* dokumen harus dilakukan persiapan terlebih dahulu sehingga dapat digunakan untuk proses utama. Proses mempersiapkan *text* atau data mentah ini disebut *text preprocessing*. *Text preprocessing* berfungsi mengubah data yang tidak terstruktur menjadi terstruktur Tu, (2019) . Proses yang dilakukan dalam preprocessing adalah sebagai berikut :

- a. **Case folding**, merupakan proses merubah kalimat data teks menjadi seragam
- b. **Cleansing**, yaitu proses membersihkan dokumen dan menyeleksi kata yang tidak diperlukan seperti html, emoticon, hashtag, mention dan url.
- c. **Stopwords**, yaitu menghilangkan kata yang kurang efektif
- d. **Steaming**, yaitu proses untuk menyaring kata yang terdapat kata sambung, kata ganti, kata depan, menjadi kata dasar dengan menghilangkan awalan atau akhiran.
- e. **Tokenization**, merupakan proses seleksi pemotongan kata dalam kalimat. Diberikan pemisah seperti tanda koma (,), titik (.), dan tanda pemisah lainnya.

Kegiatan penelitian merupakan suatu proses memperoleh atau mendapatkan suatu pengetahuan atau memecahkan permasalahan yang dihadapi, yang dilakukan secara ilmiah, sistematis dan logis[5]–[7]. Dalam penelitian di bidang apapun, tahapan-tahapan itu pada umumnya memiliki kesamaan, walaupun ada beberapa hal sering terjadi pemodifikasian dalam pelaksanaannya oleh peneliti sesuai dengan kondisi dan situasi yang dihadapi tanpa mengabaikan prinsip-prinsip umum yang digunakan dalam proses penelitian[8]–[11]. Tahapan penelitian yang dilakukan dapat dilihat pada Gambar 1 berikut.





Gambar 1. Tahapan Penelitian

3. HASIL DAN PEMBAHASAN

Contoh dari tahap pengambilan data dari twitter di hasilkan data sebagai berikut :

Tabel 1. Data tweet

No	Tweet
1	partai Golkar dan Demokrat akan Bertanding pada saat kampanye 2009
2	Pertandingan pertama antara Persema vs Persebaya malang
3	Sangat besar hapan wasit saay tanding sepakbola dpt berlaku adil
4	partai demokrat menang 2019 krna ada figur sby
5	Pertandingan sepakbola persebaya pada kampanye pemilu 2009 akan ditunda

Setelah melawati tahapan *preprocessing*, akan di dapat data sebagai berikut :

Tabel 2. Hasil Tahapan *Preprocessing*

No	Tweet	Hasil	Sentimen
1	partai Golkar dan Demokrat akan Bertanding pada saat kampanye 2009	partai golkar demokrat tanding kampanye 2009	Negatif
2	Pertandingan pertama antara Persema vs Persebaya malang	tanding pertama persema persebaya malang	positif
3	Sangat besar hapan wasit saay tanding sepakbola dpt berlaku adil	besar wasit tanding sepakbola adil	positif
4	partai demokrat menang 2019 krna ada figur sby	partai demokrat menang 2019 figur sby	negatif
5	Pertandingan sepakbola persebaya pada kampanye pemilu 2009 akan ditunda	tanding sepakbola persebaya kampanye pemilu 2009 tunda	?

Selanjutnya setelah mendapatkan data yang sudah melalui tahap preprocessing akan dihitung bobotnya dengan TF-IDF berikut hasilnya :

Tabel 3. Hasil Perhitungan TF-IDF

term	tf					DF	idf	wdf=tf.idf					
	D1	D2	D3	D4	D5			D1	D2	D3	D4	D5	
partai	1			1		2	0,39794	0,39794	0	0	0,39794		0



golkar	1			1	0,69897	0,69897	0	0	0	0
demokrat	1		1	2	0,39794	0,39794	0	0	0,39794	0
tanding	1	1	1	4	0,09691	0,09691	0,09691	0,09691	0	0,09691
kampanye	1		1	2	0,39794	0,39794	0	0	0	0,39794
2009	1		1	3	0,22185	0,221849	0	0	0,221849	0,221849
pertama	1			1	0,69897	0	0,69897	0	0	0
persema	1			1	0,69897	0	0,69897	0	0	0
persebaya	1		1	2	0,39794	0	0,39794	0	0	0,39794
malang	1			1	0,69897	0	0,69897	0	0	0
besar		1		1	0,69897	0	0	0,69897	0	0
wasit		1		1	0,69897	0	0	0,69897	0	0
sepakbola		1	1	2	0,39794	0	0	0,39794	0	0,39794
adil		1		1	0,69897	0	0	0,69897	0	0
menang			1	1	0,69897	0	0	0	0,69897	0
pemilu		1	1	2	0,39794	0	0	0	0,39794	0,39794
figur		1		1	0,69897	0	0	0	0,69897	0
sby		1		1	0,69897	0	0	0	0,69897	0
tunda			1	1	0,69897	0	0	0	0	0,69897

Selanjutnya adalah menghitung kemiripan antar dokumen dengan menggunakan perhitungan *cosine similarity* karena yang akan dicari adalah nilai sentimen dari data tweet terakhir, maka data tweet (D5) dihitung *similaritynya* dengan semua data yang ada hasilnya adalah sebagai berikut.

Tabel 4. Hasil Perhitungan TF-IDF Lanjutan

wd5*wdi			
D1	D2	D3	D4
0	0	0	0
0	0	0	0
0	0	0	0
0,009391551	0,009392	0,0093916	0
0,158356251	0	0	0
0,049216868	0	0	0,049217
0	0	0	0
0	0	0	0
0	0,158356	0	0
0	0	0	0
0	0	0	0
0	0	0	0
0	0	0,1583563	0
0	0	0	0
0	0	0	0
0	0	0	0,158356
0	0	0	0
0	0	0	0
0,21696	0,16775	0,16775	0,20757





Langkah selanjutnya hitung panjang setiap dokumen termasuk D5 dengan cara Kuadratkan bobot setiap term dalam setiap dokumen, jumlahkan nilai kuadrat tersebut dan kemudian akarkan hasil dari dokumen yang sudah di kuadratkan seperti contoh pada tabel dibawah ini.

Tabel 5. Hasil Perhitungan TF-IDF Lanjutan

Panjang Vektor				
D1	D2	D3	D4	D5
0,158356	0	0	0,158356	0
0,488559	0	0	0	0
0,158356	0	0	0,158356	0
0,009392	0,009392	0,009392	0	0,009392
0,158356	0	0	0	0,158356
0,049217	0	0	0,049217	0,049217
0	0,488559	0	0	0
0	0,488559	0	0	0
0	0,158356	0	0	0,158356
0	0,488559	0	0	0
0	0	0,488559	0	0
0	0	0,488559	0	0
0	0	0,158356	0	0,158356
0	0	0,488559	0	0
0	0	0	0,488559	0
0	0	0	0,158356	0,158356
0	0	0	0,488559	0
0	0	0	0,488559	0
0	0	0	0	0,488559
1,022236	1,633425	1,633425	1,989963	1,180592
1,01106	1,27806	1,27806	1,41066	1,08655

Lalu langkah selanjutnya mencari kemiripan antar dokumen, hitung kemiripan D1 Dengan D5, D2 dengan D5, D3 dengan D5 dan D4 dengan D5 sebagai berikut :

Tabel 6. Menghitung Kemiripan Dokumen

Dokumen	Rumus	Hasil
COS(D5,D1)	$0,26196/(1,08655*1,01106)$	0,1975
COS(D5,D2)	$0,16775/(1,08655*1,27806)$	0,1208
COS(D5,D3)	$0,16775/(1,08655*1,27806)$	0,1208
COS(D5,D4)	$0,20757/(1,08655*1,41066)$	0,13542

Lalu hasil pada Tabel 6 Diatas diurutkan berdasarkan jarak dengan nilai terbesar ke terkecil, sehingga dapat dilihat pada Tabel 7:

Tabel 7. Menghitung Kemiripan Dokumen lanjutan

D1	D2	D3	D4
0,1975	0,1208	0,1208	0,13542



Lalu hasil pada tabel 6 Diatas diurutkan berdasarkan jarak dengan nilai terbesar ke terkecil, sehingga menjadi :

Tabel 8. Urutan jarak dari dokumen

1	2	3	4
D1	D4	D2	D3
0,1975	0,13542	0,1208	0,1208

Langkah selanjutnya menentukan kelas dari D5 dengan mengambil dari K sebanyak K (K=3) yang paling tinggi tingkat kemiripannya dengan D1 maka hasilnya sebagai berikut

Tabel 9. Menentukan Kelas dari D1

D1	D4	D2
Negatif	Negatif	Positif

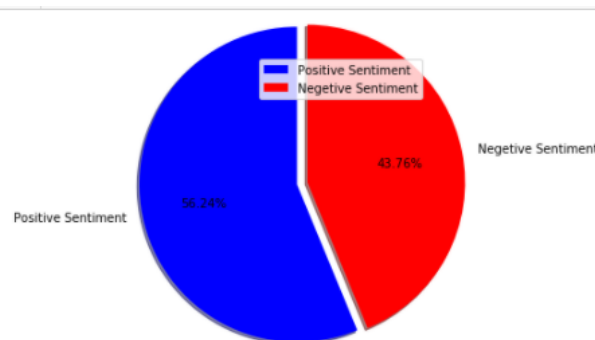
Pilih kelas yang paling banyak kemunculannya. Untuk K=3 :

Kelas positif, diwakili oleh 2 dokumen yaitu D1 dan D4.

Kelas negatif, hanya diwakili oleh D2.

Kesimpulan D5 terklarifikasi ke Kelas **Negatif**.

Dari hasil pengujian perbandingan antara tweet yang di tuliskan oleh masyarakat terkait Pembelajaran daring yaitu tweet yang bersifat positif memiliki presentase mencapai 56,24% sedangkan tweet yang bersifat negatif 43,76% dengan total data 1825, dan perbandingan kelas positif mencapai angka 1039 sedangkan kelas negatif mencapai 806 data.



Gambar 2. Persentase Perbandingan Kelas Negatif Dan Positif

Berdasarkan penelitian yang telah dilakukan, dapat diketahui bahwa sebanyak 56,24% tweet positif sedangkan tweet yang bersifat negatif 43,76% dengan total data 1825.

Hasil pengujian menggunakan confusion matrix dapat dilihat pada tabel 10 berikut ini.

Tabel 10. Hasil Pengujian Sistem Dengan *Confusion Matrix*

Pengujian	Hasil
<i>Accuration</i>	85%
<i>Precision</i>	87%
<i>Recall</i>	86%
<i>F measure</i>	87%
<i>Error Rate</i>	0.12%



Algoritma K Nearest Neighbor dengan pembobotan kata TF-IDF, dengan jumlah dataset 1825 data. Dataset dibagi menjadi dua data yaitu 80% data training dan 20% data testing, menghasilkan tingkat akurasi sebesar yang sebesar 84,93% pada pengujian K=10. Sedangkan presisi mencapai 87%, *recall* 87%, *f measure* 87% serta memiliki error rate 0,12%.

4. KESIMPULAN

Dari hasil pengujian perbandingan antara tweet yang di tuliskan oleh masyarakat terkait Pembelajaran daring yaitu tweet yang bersifat positif memiliki presentase mencapai 56,24% sedangkan tweet yang bersifat negatif 43,76% dengan total data 1825, dan perbandingan kelas positif mencapai angka 1039 sedangkan kelas negatif mencapai 806 data. Hasil pengujian menggunakan *confusion matrix* dari Algoritma K Nearest Neighbor dengan pembobotan kata TF-IDF, dengan jumlah dataset 1825 data. Dataset dibagi menjadi dua data yaitu 80% data training dan 20% data testing, menghasilkan tingkat akurasi sebesar yang sebesar 84,93% pada pengujian K=10. Sedangkan presisi mencapai 87%, *recall* 87%, *f measure* 87% serta memiliki error rate 0,12%.

5. REFERENCES

- [1] A. P. Opute, B. O. Irene, and C. G. Iwu, "Tourism service and digital technologies: A value creation perspective," *African J. Hosp. Tour. Leis.*, 2020.
- [2] Y. Mejova, "Sentiment Analysis: An Overview Comprehensive Exam Paper," *Science (80-)*, 2009.
- [3] B. Trstenjak, S. Mikac, and D. Donko, "KNN with TF-IDF based framework for text categorization," *Procedia Eng.*, vol. 69, pp. 1356–1364, 2014.
- [4] Y. Tu, "Machine learning," in *EEG Signal Processing and Feature Extraction*, 2019. doi: 10.1007/978-981-13-9113-2_15.
- [5] P. Lestari, D. Darwis, and D. Damayanti, "Komparasi Metode Economic Order Quantity Dan Just In Time Terhadap Efisiensi Biaya Persediaan," *J. Akunt.*, vol. 7, no. 1, pp. 30–44, 2019.
- [6] D. A. Megawaty and D. Santia, "Assessment of The Alignment Maturity Level of Business and Information Technology at CV Jaya Technology," in *2019 International Conference on Computer Science, Information Technology, and Electrical Engineering (ICOMITEE)*, 2019, pp. 54–58.
- [7] S. Setiawansyah, A. T. Priandika, B. Ulum, A. D. Putra, and D. A. Megawaty, "UMKM Class Determination Support System Using Profile Matching," *Bull. Informatics Data Sci.*, vol. 1, no. 2, pp. 46–54, 2022.
- [8] A. T. Priandika, "SISTEM PENGENDALIAN INTERNAL MONITORING INVENTORY OBAT MENGGUNAKAN SUPPLY CHAIN MANAGEMENT," *J. Ilm. BETRIK Besemah Teknol. Inf. dan Komput.*, vol. 12, no. 1, pp. 36–44, 2021.
- [9] S. A. Widiana, S. Sintaro, R. Arundaa, E. Alfonsius, and D. Lapihu, "Aplikasi Penjualan Baju Berbasis Web (E-Commerce) dengan Formulasi Penyusunan Kode," *J. Inf. Technol. Softw. Eng. Comput. Sci.*, vol. 1, no. 1 SE-Articles, pp. 35–43, Jan. 2023, doi: 10.58602/itsecs.v1i1.11.
- [10] F. Hamidy and I. Yasin, "Implementation of Moving Average for Forecasting Inventory Data Using CodeIgniter," *J. Data Sci. Inf. Syst.*, vol. 1, no. 1, pp. 17–23, 2023.
- [11] M. N. D. Satria, "Sistem Pendukung Keputusan Penerimaan Staff Administrasi Menggunakan Metode VIKOR," *J. Artif. Intell. Technol. Inf.*, vol. 1, no. 1, pp. 39–49, 2023.

