

ANALISIS SENTIMEN MASYARAKAT INDONESIA TERHADAP PEMBELAJARAN ONLINE DARI DI MEDIA SOSIAL TWITTER MENGGUNAKAN *LEXICON* DAN *K-NEAREST NEIGHBOR*

Ahmad Halimi¹, Kusri², M. Rudyanto Arief³
^{1,2,3} Magister Teknik Informatika UNIVERSITAS AMIKOM Yogyakarta

Info Artikel	ABSTRAK
<p><u>Riwayat Artikel</u> Diterima: 27-05-2021 Disetujui: 26-06-2021</p> <p><u>Kata Kunci</u> Analisis sentimen; Lexicon; K-Nearest Neighbors; TF-IDF; Pembelajaran online;</p> <p>e-mail* ahmad.1259@students.amikom.ac.id</p>	<p>Semenjak keputusan pembelajaran secara online oleh mendikbud di masa pandemi covid-19, membuat masyarakat emosi serta meluapkan opini mereka ke media sosial. Tujuan penelitian ini untuk mengetahui sentimen positif, negatif ataupun netral terkait kebijakan mendikbud terhadap pembelajaran online yang keberlangsungan. Data yang diperoleh dari media sosial twitter yang berkaitan dengan hastag-hastag pembelajaran online. Di penelitian ini untuk melakukan klasifikasi sentimen menggunakan Lexicon dan K-Nearest Neighbors untuk mengetahui sentimen yang terjadi pada tweet dan pengujian hasil akurasi menggunakan K-Nearest Neighbors dengan validasi Nilai K=3 sampai K=9. Maka dalam pengujian data yang telah dikumpulkan kurang baik untuk dilanjutkan klasifikasi sentimennya, di perlunya preprocessing kata pada kalimat tweet supaya lebih baik serta menghapus kalimat yang kembar, sehingga proses klasifikasi sentimen dengan K-Nearest Neighbors dan Lexicon sangat baik. Hasil dari penelitian ini memperoleh tweet positif sebanyak 1433, tweet negatif sebanyak 461 dan tweet sebanyak netral 71 untuk K-Nearest Neighbors dan Lexicon tweet sebanyak positif 759, tweet negatif sebanyak 653 dan tweet netral sebanyak 553 serta hasil akurasi dari pengujian validasi nilai K yang terbaik pada nilai K=3 dengan data sentimen K-Nearest Neighbors sebesar 80.66% dan data sentimen sebesar Lexicon 80.92%. Maka hasil lebih baik pada data sentimen Lexicon dengan validasi K=3.</p>

1. PENDAHULUAN

Pembelajaran daring atau pembelajaran yang dilakukan secara online merupakan topik yang saat ini sedang populer di Indonesia, hal ini dikarenakan surat edaran yang dikeluarkan oleh Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 3 Tahun 2020 Tentang Pencegahan *Coronavirus Disease (Covid-19)* Pada Satuan Pendidikan dengan tujuan mencegah perkembangan serta penyebaran Covid-19 di lingkungan satuan pendidikan[1][2]. Banyak masyarakat yang belum siap terhadap keputusan pembelajaran daring yang harus dilaksanakan oleh Satuan Pendidikan, beberapa faktor yang menjadi kendala dalam terlaksananya pembelajaran daring yakni fasilitas serta media yang akan digunakan untuk menunjang pembelajaran daring yang membutuhkan biaya cukup mahal serta minimnya pengetahuan terhadap media yang akan digunakan seperti *zoom*, *google meet*, *google classroom* dan media lainnya[3].

Disaat kebijakan pembelajaran secara online berjalan selama satu tahun pasti menimbulkan opini-opini kalangan masyarakat yang bermacam-macam yang telah diluapkan ke media sosial

salah satunya twitter [4]. Twitter dikalangan masyarakat sering kali digunakan sebagai curhat hati terhadap suatu hal unik yang harus dibicarakan oleh pengguna lainnya untuk mengetahui tanggapan seperti pembelajaran online dimasa pandemi covid-19 yang begitu ramai di perbincangkan [5]. Maka dibutuhkan proses klasifikasi sentimen terhadap pembelajaran secara online di twitter untuk mengali sebuah informasi sentimen yang di dapatkan dari data tweet pengguna twitter [6][7]. Akan tetapi untuk memperoleh data yang baik dibutuhkan proses *text preprocessing* supaya dataset di peroleh lebih bersih dari url, gambar dan simbol-simbol tidak dibutuhkan, Serta dilakukan klasifikasi sentimen dengan algoritma *K-Nearest Neighbors* dan *Lexicon* [8].

Penelitian sebelumnya juga menggunakan metode klasifikasi *K-Nearest Neighbors* dan *Lexicon Based* untuk menentukan sentimen yang dihasilkan dari tweet pengguna media sosial twitter terhadap dompet digital. tujuan penelitian ini untuk melakukan visualisasi sentimen tentang dompet elektronik seperti GO PAY, OVO, LinkAja dari opini pengguna twitter dengan melakukan klasifikasi data menggunakan *K-Nearest Neighbors* dan *Lexicon based*. Pada penelitian terkait sumber data yang diperoleh dari penarikan data melalui API twitter, serta dilanjutkan dengan *preprocessing* untuk membersihkan data dari kata tidak baku yang akan proses kembali untuk mengetahui sentimen positif atau negatif dari *Lexicon Based*. Pada data tweet yang telah dikelola, maka dibutuhkan sentuhan terakhir sebagai klasifikasi dengan *K-Nearest Neighbors* untuk mendapatkan nilai akurasi yang terbaik dari dompet elektronik tersebut. Hasil akhir dari metode *K-Nearest Neighbors* pada data GOPAY sebanyak 337 tweet, OVO sebanyak 357 tweet, Link Aja sebanyak 255 tweet dan dilakukan perhitungan confusion matrix, Maka Menghasilkan nilai akurasi pada data GOPAY sebanyak 94.05% positif serta negatif sebanyak 5.95%, OVO sebanyak 86.91% positif serta negatif 13.09%, dan terakhir untuk LinkAja sebanyak 76.31% serta negatif sebanyak 23.69% [9].

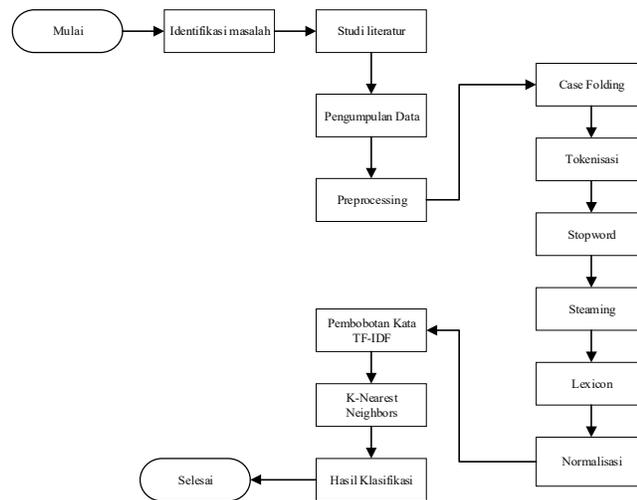
K-nearest neighbors mempunyai cara kerja untuk mengklasifikasi terhadap kondisi objek yang memiliki *variabel K* berlandaskan dari *data testing* dengan kondisi kedekatan jarak pada objek *data training*, serta ketentuan pemberian nilai *variabel K* jangan sampai melebihi dari *data training* ditentukan dan untuk nilai *K* harus bernilai ganjil atau dan tidak boleh kosong[10]. Tujuan dari algoritma tersebut untuk mengklasifikasi sebuah objek berlandaskan *variabel K* dengan tetangga terdekat sebagai *query instance* untuk nilai prediksi yang baru[11]. Kekurangan dari *K-nearest neighbors* salah satunya metode yang digunakan keseluruhan fitur perhitungan untuk besar jarak yang mengakibatkan beratnya komutasi pada *K-nearest neighbors*, maka untuk mengatasi permasalahan tersebut butuh metode *Term Frequency – Inverse Document Frequency (TF-IDF)*[12].

Metode *Lexicon* merupakan salah satu metode untuk melakukan klasifikasi pada kalimat yang mengandung opini atau tidak dari kecocokan kata pada dataset *Lexicon* yang berbahasa indonesia, serta pada kata tersebut akan di memiliki nilai *polarity* untuk mengetahui sentimen apa yang terjadi pada kalimat tersebut. Sehingga mempermudah dalam penentuan klasifikasi pada kalimat yang bermakna sentimen positif, negatif atau netral [13].

Oleh sebab itu, penelitian ini bertujuan untuk mengetahui hasil akurasi dari algoritma *K-Nearest Neighbors* dari perbandingan klasifikasi sentimen *K-Nearest Neighbors* dan *Lexicon* dengan *confusion matrix* yang digunakan sebagai analisis sentimen masyarakat terhadap opini positif, negatif, ataupun netral berkaitan tentang kebijakan pemerintah terhadap pembelajaran secara online. Maka dalam proses klasifikasi dengan *K-Nearest Neighbors*, dibutuhkan data tweet yang telah di peroleh dari tanggal 1 januari 2021 sampai 26 april 2021 dengan tweet yang berbahasa indonesia yang berbasis teks serta bisa dijadikan sebagai acuan evaluasi terhadap pembelajaran secara online di masa depan

2. METODE PENELITIAN

Alur penelitian yang akan dijelaskan dalam bentuk skema gambar penelitian untuk memperjelas peneliti lakukan, sebagai berikut:



Gambar 1. Alur Penelitian

Pada Gambar 1 menjelaskan alur peneliti yang dimulai dari indentifikasi masalah yang di angkat pada penelitian ini. Studi literatur tujuan mencari referensi dari jurnal, buku dan lain-lainnya yang berkaitan penelitian yang dilakukan. Preprocessing untuk melakukan perbaikan kata serta membuang kata yang tidak bermakna, supaya waktu klasifikasi lebih baik. Pada preprocessing memiliki ditahap proses yaitu *Case Folding*, *Tokenisasi*, *Stopword*, *Steaming*, *Lexicon*, *Normalisasi*. Pembobotan kata TF-IDF untuk mengubah kata menjadi nilai yang akan di proses ke klasifikasi *K-Nearest Neighbors*. *K-Nearest Neighbors* untuk klasifikasi sentimen dan menentukan hasil akurasi dari data sentimen sudah proses *Lexicon* dan *K-Nearest Neighbors*. Hasil klasifikasi akan di sajikan pada *confusion matrix* untuk membandingkan hasil akurasi pada data sentimen *Lexion* dan *K-Nearest Neighbors*. Berikut merupakan penjelasan secara detail:

1. Identifikasi masalah

Pada tahap awal peneliti lakukan yaitu mengidentifikasi permasalahan di penelitian yang dibuat tentang sentimen terhadap pembelajaran online di kalangan masyarakat indonesia yang masih belum dikatakan efektif sebagai kalangan masyarakat, sehingga perlunya analisis sentimen terhadap opini tersebut untuk dijadikan sebuah informasi relevan terhadap pembelajaran online.

2. Studi literatur

Pada studi literatur melakukan kajian terhadap teori yang relevan pada penelitian sejenis yang sudah dilakukan penelitian sebelumnya, karena tujuan dari penelitian ini mencari suatu permasalahan yang ada serta diberi hasil kesimpulan sebagai solusi untuk peneliti lakukan sehingga mendapat sebuah hasil yang memuaskan. beberapa teori yang digunakan sebagai teori tentang opini, seperti data mining, machine learning, analisis sentimen serta *K-Nearest Neighbors*.

3. Pengumpulan data

Pengumpulan data yang dilakukan pertama dengan mendaftarkan akun twitter terlebih dahulu, kemudian daftar sebagai seorang developer atau pengembang aplikasi di situs developer.twitter.com untuk mendapatkan akses api tweet yang nantinya akan digunakan sebagai penarikan data tweet pengguna. Penarikan data tweet tersebut dibutuhkan kata kunci seperti "#PJJ", "#Belajardirumah", "#BelajarOnline", "Belajar Daring", dan "Daring" serta ketentuan tweet harus berbahasa indonesia dan pengumpulan data dimulai dari tanggal 1 januari 2021 sampai 26 april 2021.

4. Preprocessing

Preprocessing merupakan proses pembersihan data yang tidak sempurna serta data yang tidak jelas untuk kelolah. mengenai langkah-langkah yang harus dilakukan dalam proses *preprocessing* tersebut sebagai berikut[14]:

a. Case folding

Case folding ini dilakukan untuk memperbaiki atau memperbarui data kata yang besar menjadi kata yang kecil (*lowercase*) dan membersihkan kata pada kalimat yang berunsur url(<http://>), add (@), hashtag atau tag pagar (#), dan delimiter seperti titik koma (,) dan tanda baca lainnya, contohnya seperti di Tabel 1.

Tabel 1. Case Folding

Tweet	Case Folding
Kenapa kuliah daring terasa lebih melelahkan dan lebih berat? Mungkin krn rumah yang tlg ditetapkan oleh kepala kita sebagai tempat bersantai dan beristirahat dijadikan tempat serius dan fokus, ya capek mental sih jadinya.	kenapa kuliah daring terasa lebih melelahkan dan lebih berat mungkin krn rumah yang tlg ditetapkan oleh kepala kita sebagai tempat bersantai dan beristirahat dijadikan tempat serius dan fokus ya capek mental sih jadinya

Pada Tabel 1 menjelaskan perbedaan hasil dari proses *case folding* untuk mengubah kata yang besar menjadi kecil, seperti pada tulisan berwarna kuning.

b. Tokenisasi

Tokenisasi ini dilakukan untuk memecah kalimat menjadi per kata, contohnya seperti di Table 2.

Tabel 2. Tokenisasi

Tweet	Tokenisasi
Kenapa kuliah daring terasa lebih melelahkan dan lebih berat? Mungkin krn rumah yang tlg ditetapkan oleh kepala kita sebagai tempat bersantai dan beristirahat dijadikan tempat serius dan fokus, ya capek mental sih jadinya.	['kenapa', 'kuliah', 'daring', 'terasa', 'lebih', 'melelahkan', 'dan', 'lebih', 'berat', 'mungkin', 'krn', 'rumah', 'yang', 'tlg', 'ditetapkan', 'oleh', 'kepala', 'kita', 'sebagai', 'tempat', 'bersantai', 'dan', 'beristirahat', 'dijadikan', 'tempat', 'serius', 'dan', 'fokus', 'ya', 'capek', 'mental', 'sih', 'jadinya']

Pada Tabel 2 menjelaskan hasil dari proses tokenisasi untuk memecah kalimat menjadi per kata dengan simbol mengurung serta diberi petik satu sebagai penanda per-kata dan dilanjutkan dengan koma sebagai lanjut kata.

c. Stopword

Stopword ini dilakukan untuk menghapus kata yang tidak anggap penting atau kata tidak bermakna, contohnya seperti di Tabel 3.

Tabel 3. Stopword

Tweet	Stopword
Kenapa kuliah daring terasa lebih melelahkan dan lebih berat? Mungkin krn rumah yang tlg ditetapkan oleh kepala kita sebagai tempat bersantai dan beristirahat dijadikan tempat serius dan fokus, ya capek mental sih jadinya.	['kuliah', 'daring', 'melelahkan', 'berat', 'rumah', 'ditetapkan', 'kepala', 'bersantai', 'beristirahat', 'dijadikan', 'serius', 'fokus', 'capek', 'mental']

Pada Tabel 3 menjelaskan hasil dari proses *stopword* untuk menghapus kata tidak penting atau tidak bermakna seperti pada kata memberi warna kuning.

d. Streaming

Streaming merupakan proses untuk melakukan perubahan semua kata pada kalimat yang memiliki imbuhan yang akan dijadikan kata dasar. kata imbuhan yang dibuang seperti kata imbuhan awalan, sisipan, akhiran, dan gabungan awalan dan akhiran, contohnya seperti di Tabel 4.

Tabel 4. Streaming

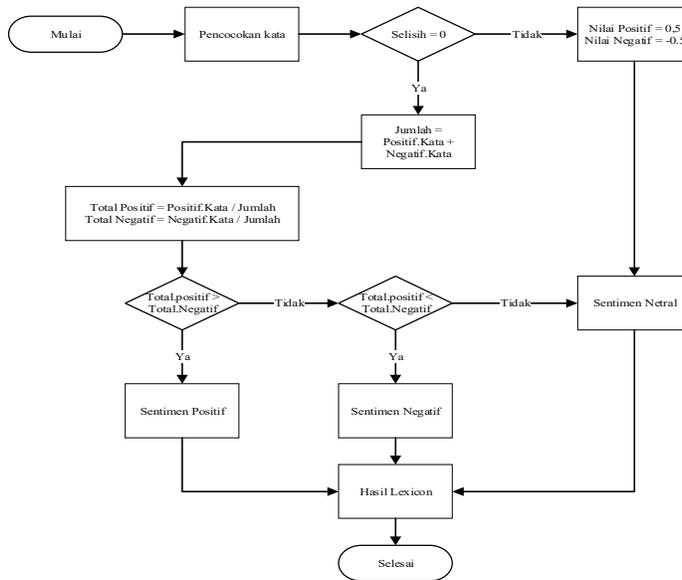
Tweet	Streaming
Kenapa kuliah daring terasa lebih melelahkan dan lebih berat? Mungkin krn rumah yang tlg ditetapkan oleh kepala kita sebagai tempat bersantai dan beristirahat dijadikan tempat serius dan fokus, ya capek mental sih jadinya.	['kenapa', 'kuliah', 'daring', 'asa', 'lebih', 'lelah', 'dan', 'lebih', 'berat', 'mungkin', 'krn', 'rumah', 'yang', 'tlg', 'tetap', 'oleh', 'kepala', 'kita', 'bagai', 'tempat', 'santa', 'dan', 'serius', 'fokus', 'ya', 'capek', 'mental', 'sih', 'jadinya']

	'istirahat', 'jadi', 'tempat', 'serius', 'dan', 'fokus', 'ya', 'capek', 'mental', 'sih', 'jadi'
--	---

Pada Tabel 4 menjelaskan hasil Streaming untuk menghilangkan kata imbuhan seperti pada kata yang diberi warna kuning.

e. *Lexicon*

Lexicon ini dilakukan untuk mengubah per kata pada kalimat menjadi nilai polaritas dan dihitung nilai sentimen pada kata dengan pencocokan kata pada dataset *lexicon* tersebut serta disimpulkan menjadi kalimat positif, negatif, dan netral. dataset yang digunakan <https://github.com/ahalimidev/Lexicon-Base.git> dan alur perhitungannya:



Gambar 2. Alur Perhitungan Lexicon

Pada Gambar 2 menjelaskan proses klasifikasi sentimen dengan *Lexicon* dengan lakukan proses steaming tweet dan dilakukan pencocokan kata dengan kamus yang sudah disedia. Maka perhitungan pada kata akan dicocokkan dengan nilai yang ada pada kamus dengan memberi nilai positif 0,5, negatif - 0,5 dan netral 0 serta pada akhir proses akan dijumlah hasil sentimen dengan *Lexicon* di setiap kalimat.

f. Normalisasi

Normalisasi ini dilakukan untuk memperbaiki kata yang tidak sempurna atau kata yang noise menjadi kata yang baku. kata tika baku yang maksud dengan kata alay, kata singkat dan tidak beracun kamus besar bahasa indonesia (KBBI), contohnya seperti di Tabel 5.

Tabel 5. Normalisasi

Tweet	Normalisasi
Kenapa kuliah daring terasa lebih melelahkan dan lebih berat? Mungkin krm rumah yang tlh ditetapkan oleh kepala kita sebagai tempat bersantai dan beristirahat dijadikan tempat serius dan fokus, ya capek mental sih jadinya.	['kenapa','kuliah', 'daring', 'melelahkan', 'dan', 'lebih', 'berat', 'mungkin', ' karena ', 'rumah', ' telah ', 'ditetapkan', 'kepala', 'oleh', 'kepala', 'kita', 'sebagai', 'tempat', 'bersantai', 'dan', 'beristirahat', 'dijadikan', 'serius', 'fokus', ' telah ', 'mental']

Pada tabel 5 menjelaskan prose normalisasi kata yang tidak baku menjadi baku, seperti di kata yang diberi warna kuning.

5. Pembobotan Kata TF-IDF

Pembobotan kata dengan TF-IDF ini merupakan proses perubahan suatu kata menjadi bentuk nilai yang akan di proses pembobotan katanya. TF (*Term Frequency*) merupakan perhitungan kata yang kemunculan ada disetiap dokumen dengan frequency terbanyak. IDF (*Inverse Document Frequency*) merupakan perhitungan kemunculan kata yang jarak pada

dokumen yang sedikit atau bisa berbanding terbalik dengan TF, maka kata yang kemunculan sedikit akan bernilai sangat tinggi. Rumus Pembobotan kata TF-IDF[15]:

$$W^{(y,d)} = TF^{(y,d)} \times IDF^{(y)} = TF^{(y,d)} \times \frac{1}{n} \quad (1)$$

$W^{(y,d)}$ = Pembobotan kata
 $TF^{(y,d)}$ = Jumlah Frequency Kata
 $IDF^{(y)}$ = Jumlah Inverse Document Frequency Setiap kata
 $DF^{(y)}$ = Jumlah Frequency Document Setiap kata
 N = Jumlah banyaknya dokumen

6. *K-nearest neighbors*

K-nearest neighbors ialah model klasifikasi yang memiliki konsep *supervised learning* yang dipakai sebagai klasifikasi suatu object yang bersumber pada *variabel* dan data *training*. Untuk melakukan klasifikasi dibutuhkan data testing untuk hitung jaraknya paling terdekat dengan objek *data training*. sehingga perhitungan antara untuk data baru menggunakan euclidean dengan jarak *data training*. Rumus perhitungan *euclidean distance*[16]:

$$Q^{(A,B)} = \sqrt{\sum_{k=1}^N (A_k - B_k)^2} \quad (2)$$

Q = Jarak data dari dua variabel A dan B
 A = *Data Training*
 B = *Data Testing*
 N = Dimensi data
 K = *Variabel data*

7. Hasil Klasifikasi

Setelah proses pengujian klasifikasi pada *K-Nearest Neighbors* selesai, selanjutnya dilakukan perhitungan akurasi dari beberapa kali pengujian pada algoritma *K-Nearest Neighbors* dengan dibantu menggunakan *confusion matrix* untuk membandingkan jumlah pengujian serta berapa tingkat akurasi dimiliki. Rumus perhitungan *confusion matrix* [17]:

$$\text{Akurasi} = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (3)$$

$$\text{Presisi} = \frac{TP}{FP + TP} \times 100\% \quad (4)$$

$$\text{Recall} = \frac{TP}{FN + TP} \times 100\% \quad (5)$$

Pada TP, dan TN menjelaskan tentang prediksi *true positive* yang positif benar serta *true negative* yang negatif benar. FP dan FN menjelaskan sebagai *false positive* yang positif salah serta FN yang Negatif yang salah. Dari *confusion matrix* maka dilakukan perhitungan nilai akurasi, presisi dan recall pad TP, TN, FP dan FN [18]. Untuk melakukan perhitungan pada akurasi, presisi dan recall akan diuji coba menggunakan perhitungan manual sebagai berikut [19]:

Tabel 6. Perhitungan Manual

Akurasi	Persisi	Recall
= (TP+TN) / (TP+FP+FN+TN)	= (TP) / (TP + FP)	= TP / (TP + FN)
= (125+5) / (125+20+25+5)	= 125 / (125 + 20)	= 125 / (125+25)
= 0.742	= 0.86	= 0.83
= 0.742 * 100 % = 74.2 %	= 0.86 * 100% = 86%	= 0.83 * 100% = 83%

3. HASIL DAN PEMBAHASAN

1. Pengumpulan Data

Tweet yang telah diambil menggunakan melalui API Twitter developer dan dilakukan programan berbasis python pada anaconda jupyter serta menggunakan plugin bantuan yaitu tweepy sebagai twitter scraper. Data yang telah diambil sebanyak 2484 tweet dari

tanggal 1 januari 2021 sampai 26 april 2021 dengan kata kunci "#PJJ", "#Belajardirumah", "#BelajarOnline", "Belajar Daring", dan "Daring". Berikut merupakan isi dari data tweet tersebut berupa 3 tweet yang ditampilkan sebagai berikut.

Tabel 7. Data Tweet

Label	Tweet
1	Siswa di Zona Oranye dan Zona Merah di Pekanbaru Belajar Daring Lagi, Belajar di Sekolah Dihentikan https://t.co/6GnLtazqRG
2	@bluetomyam benerrrr, masa masuk masuk disuruh ujian sekolah jadi satu minggu ini bener bener harus ekstra belajarðŸˆš soalnya gapaham semua materi waktu daring
3	@flamingkow @hheadbanging @Elenaskyee @collegemenfess Kalau mau online banyak hal yang harus diperbaiki dulu. Situasi, motivasi belajar, teknis dan sistem pembelajaranya dan jaringan itu yang paling jadi dasar fasilitator dalam sistem daring yang harus menjangkau semuanya

Pada Tabel 7 ini masih belum dikatakan dataset yang baik, karena masih ada kata-kata yang tidak sempurna untuk melakukan pengujian di analisis sentimen. Dibutuhkan preprocessing data untuk membersihkan data dari kata tidak sempurna.

2. Preprocessing

Setelah diperoleh hasil pengambilan data melalui twitter scraping, maka dilanjutkan dengan proses preprocessing untuk memperbaiki kata dan membuang kata yang tidak berguna dan membuang data yang kembar atau sama. Langkah-langkah dilakukan untuk *preprocessing* data: *case folding*, *tokenisasi*, *stopword*, *stemming*, dan normalisasi serta menghasilkan data bersih keseluruhan dataset menjadi 1965 tweet. Berikut merupakan isi dari data *preprocessing* tersebut berupa 3 data ditampilkan.

Tabel 8. Preprocessing

Label	Tweet
1	siswa zona oranye zona merah pekanbaru belajar daring belajar sekolah henti
2	benar masuk masuk suruh ujian sekolah minggu benar benar ekstra belajar tidak paham materi daring
3	Online perbaiki situasi motivasi belajar teknis sistem pembelajaran jaringan dasar fasilitator sistem daring jangkau

Pada Tabel 8 setelah melalui *prepreproseccing* tidak ditemukan sebuah kata yang tidak sempurna seperti simbol, tanda baca, dan kata tidak formal. Selanjutnya dilakukan proses *Lexicon* untuk mengetahui nilai *porality* dan sentimen dari tweet telah *preprocessing* di tahapan *steaming*. Berikut merupakan hasil dari *Lexicon* yang ditampilkan dalam bentuk 3 data.

Tabel 9. Lexicon

Label	Tweet	Polarity	Sentimen
1	siswa zona oranye zona merah pekanbaru belajar daring belajar sekolah henti	13	positif
2	benar masuk masuk suruh ujian sekolah minggu benar benar ekstra belajar tidak paham materi daring	0	netral
3	Online perbaiki situasi motivasi belajar teknis sistem pembelajaran jaringan dasar fasilitator sistem daring jangkau	-6	negatif

Pada Tabel 9 tentang hasil nilai *polarity* dan sentimen hanya beberapa tweet yang ditampilkan, Tetapi hasil sebenarnya dari keseluruhan data tweet telah *preprocessing* sampai ke *Lexicon* dengan jumlah sentimen positif sebanyak 759 tweet, negatif sebanyak 653 tweet dan netral sebanyak 553 tweet dari keseluruhan dataset 1965 tweet.

3. TF-IDF

Tahap selanjutnya dilakukan perhitungan pembobotan kata pada dataset telah di *preprocessing* data tersebut, sehingga siap untuk dihitung pembobotan kata dengan TF-IDF serta dibagi menjadi beberapa dokumen (D1, D2, D3, D4, D5, D6) yang disesuaikan dengan dataset yang dimiliki saat ini.

Tabel 10. Term ferm frequency (ft) dan document frequency(df)

Term	Term Frequency (tf)			df
	d1	d2	d3	
Sekolah	1	1	1	3

Belajar	1	1	0	2
Tugas	0	0	0	0
Ujian	0	1	0	1
Online	0	0	1	1
Daring	1	0	1	2

Pada tabel 10 menghitung beberapa banyak term(kata) yang muncul pada setiap dokumen 1 sampai 3, sehingga dalam perhitungan term(kata) dapat diberi nilai satu atau jika melebihi bisa tambah satu dengan kondisi kemunculan kata.

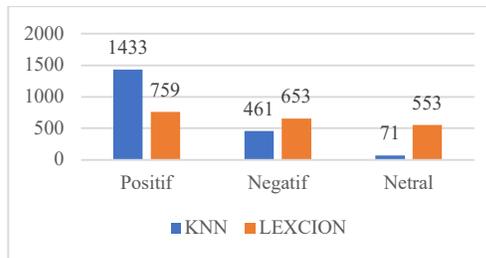
Tabel 11. Term inverse frequency (idf) dan pembobotan kata

idf	Weighting = tf.idf		
	d1	d2	d3
0,176091	0,176091	0,176091	0,176091
0,176091	0,176091	0,176091	0
0,778151	0	0	0
0,778151	0	0	0
0,778151	0	0	0
0	0	0	0

Pada table 11 menjelaskan perhitungan dari pembobotan kata dengan hitung dari banyak kata yang muncul dari setiap dokumen 1 sampai 3 dengan menggunakan rumus $\log=N/df$, dimana kata N pada rumus menjelaskan jumlah keseluruhan data yang dimiliki setiap dokumen.

4. *K-Nearest Neighbors*

Pada proses klasifikasi menggunakan *K-Nearest Neighbors* untuk melakukan prediksi klasifikasi sentimen dari data sentimen *Lexicon*, Maka untuk mengetahui hasil sentimen yang diperoleh dari *K-Nearest Neighbors* dibutuhkan proses pembobotan kata tweet dengan membandingkan hasil sentimen lexicon pada proses klasifikasi *K-Nearest Neighbors*. Berikut hasil perbandingan klasifikasi sentimen dengan *Lexion* dan *K-Nearest Neighbors*:



Gambar 3. Perbandingan Klasifikasi Sentimen

Pada gambar 3 menjelaskan hasil perbandingan data sentimen menggunakan *Lexicon* dan *K-Nearest Neighbors* yang menghasilkan sentimen positif, negatif, ataupun netral dari proses klasifikasi sentimen tersebut. Sehingga memperoleh hasil sentimen yang berbeda dari *Lexicon* yaitu positif 759 tweet, negatif 653 tweet, dan netral 553 tweet, serta dari *K-Nearest Neighbors* menghasilkan sentimen positif 1433 tweet, negatif 461 tweet, dan netral 71 tweet.

Tabel 12. Hasil pengujian prediksi sentimen

Tweet	Lexicon	K-NN
Diliburkan (belajar dirumah via daring)	positif	positif
Kasus Covid-19 Meningkat, Proses Belajar Mengajar di Anambas Dilakukan Secara Daring https://t.co/qsasIaGS2b	positif	negatif
@kuromibo0 tbh pjok pas belajar daring ini agak useless HSJSK 🤔 please jangan hujat	positif	negatif

Pada Tabel 12 menghasilkan proses prekdisi klasifikasi *K-Nearest Neighbors* yang menghasilkan perbedaan sentimen dari *Lexicon* dan *K-Nearest Neighbors* terkait tweet terhadap pembelajaran secara online di masa pandemi covid-19.

5. Hasil Klasifikasi

Untuk mendapatkan hasil akurasi dari perbandingan data sentimen yang telah di proses *Lexicon* dan *K-Nearest Neighbors* dengan cara membagi data menjadi *data training* 80% dan *data testing* 20% dari jumlah tweet 1965. Pada proses pengujian menggunakan *K-Nearest Neighbors* dibutuhkan proses validasi nilai K yang akan diuji dengan membagi K menjadi K=3, K=5, K=7 dan K=9 serta hasil akurasi yang didapatkan akan dibandingkan menggunakan *confusion matrix* dari data sentimen *Lexicon* dan *K-Nearest Neighbors*.

Tabel 13. *confusion matrix*

Nilai K	Akurasi	Persesi	Recall	Data Sentimen
K= 3	80.66	0.49	0.50	KNN
K = 5	75.32	0.43	0.45	
K= 7	74.81	0.42	0.44	
K = 9	74.05	0.39	0.43	
K= 3	80.92	0.80	0.81	Lexicon
K = 5	72.77	0.72	0.72	
K= 7	69.72	0.68	0.70	
K = 9	70.48	0.69	0.71	

Pada table 13 menjelaskan hasil akurasi dari perbandingan data sentimen digunakan dari *K-Nearest Neighbors* dan *Lexicon* dengan melakukan validasi pengujian nilai K=3 sampai K=9 menghasilkan menggunakan klasifikasi *K-Nearest Neighbors* dengan tingkat akurasi 80.66% dari nilai K=3 pada data sentimen *K-Nearest Neighbors*, Serta data sentimen *Lexicon* menghasilkan 80.92% dari pengujian nilai K=3. Maka hasil dari perbandingan tersebut yang lebih unggul ialah dari data sentimen *Lexicon* dengan tingkat akurasi yang baik dari pada data sentimen *K-Nearest Neighbors* dengan pengujian nilai K=3, Akan tetapi dari pengujian selanjutnya pada K=5 sampai K= 9 dapat mengejar hasil nilai K pada data sentimen *Lexicon*.

4. KESIMPULAN DAN SARAN

Kesimpulan dari penelitian ini adalah hasil dari jawaban pengujian akurasi serta memperoleh hasil sentimen dari *Lexicon* dan *K-Nearest Neighbors* terhadap pembelajaran secara online di masa pandemi covid-19 dari media sosial twitter dengan cara menarik data tweet sebanyak 2484. Serta dilakukan preprocessing sehingga dapat mengubah kata tidak baik menjadi baik dan membuang kata tidak dibutuhkan dalam proses klasifikasi sentimen. Setelah preprocessing dilanjutkan dengan menentukan hasil sentimen dari kalimat tweet tersebut menggunakan *Lexicon* dan dilakukan prediksi sentimen dengan *K-Nearest Neighbors*. Maka hasil sentimen yang diperoleh dari masing-masing klasifikasi yaitu tweet positif sebanyak 1433, tweet negatif sebanyak 461 dan tweet sebanyak netral 71 untuk klasifikasi KNN dan Klasifikasi *Lexicon* tweet sebanyak positif 759, tweet negatif sebanyak 653 dan tweet netral sebanyak 553.

Pada pengujian selanjutnya, dilakukan klasifikasi dengan *K-Nearest Neighbors* untuk mendapatkan hasil akurasi dari pengujian validasi nilai K=3 sampai K=9 yang ditampung *confusion matrix* dengan data sentimen *Lexicon* dan *K-Nearest Neighbors*. Maka Hasil yang diperoleh dari perbandingan data sentimen tersebut yaitu *K-Nearest Neighbors* dengan tingkat akurasi 80.66% dari nilai K=3 pada data sentimen *K-Nearest Neighbors*, Serta data sentimen *Lexicon* menghasilkan 80.92% dari pengujian nilai K=3.

Pada penelitian selanjutnya untuk topik yang sama dibutuhkan banyak dataset yang dikumpulkan berdasarkan kebijakan MENDIKBUD untuk melakukan pembelajaran secara online atau belajar dirumah saja. Serta diharapkan bisa memperbaiki data pada preprocessing yang tidak maksimal disebabkan dataset stopword dan normalisasi kurang begitu banyak datasetnya. Maka dibutuhkan dataset yang bisa mengubah kata-kata tidak baku menjadi baku berdasarkan KBBI, Sehingga proses klasifikasi tersebut lebih baik, Serta dibutuhkan optimalisasi pada *K-Nearest Neighbors* untuk meningkatkan nilai akurasi lebih baik atau bisa menggunakan algoritma yang berbeda untuk topik penelitian ini.

5. DAFTAR PUSTAKA

- [1] “Kementerian Pendidikan dan Kebudayaan » Republik Indonesia.” <https://www.kemdikbud.go.id/main/blog/2020/03/surat-edaran-pencegahan-covid19-pada-satuan-pendidikan> (accessed Jun. 09, 2021).
- [2] Humas, “Sekretariat Kabinet Republik Indonesia | Inilah Perubahan Kebijakan Pendidikan Selama Masa Pandemi Covid-19- Sekretariat Kabinet Republik Indonesia,” *Https://Setkab.Go.Id/*. 2020.
- [3] D. I. Susanti and J. Y. Prameswari, “Adaptasi Blended Learning di Masa Pandemi COVID-19 untuk Pembelajaran Bahasa Inggris di Sekolah Dasar,” *Ling. Susastra*, vol. 1, no. 2, 2020, doi: 10.24036/ls.v1i2.8.
- [4] A. K. Fauziyyah, “ANALISIS SENTIMEN PANDEMI COVID19 PADA STREAMING TWITTER DENGAN TEXT MINING PYTHON,” *J. Ilm. SINUS*, vol. 18, no. 2, 2020, doi: 10.30646/sinus.v18i2.491.
- [5] M. Syarifuddin, “Analisis Sentimen Opini Publik Mengenai Covid-19 Pada Twitter Menggunakan Metode Naïve Bayes Dan Knn,” *Inti Nusa Mandiri*, vol. 15, no. 1, 2020.
- [6] R. Taufan, “Sentimen Analisis Twitter Terhadap Isolasi Diri Masyarakat Indonesia Akibat Dampak Covid-19,” *MATICS*, vol. 12, no. 2, 2021, doi: 10.18860/mat.v12i2.9329.
- [7] M. D. Alizah, A. Nugroho, U. Radiyah, and W. Gata, “Sentimen Analisis Terkait Lockdown pada Sosial Media Twitter,” *Indones. J. Softw. Eng.*, vol. 6, no. 2, 2020, doi: 10.31294/ijse.v6i2.8991.
- [8] M. M. Rizki, “Analisis sentimen terhadap produk otomotif dari twitter menggunakan kombinasi algoritma k-nearest neighbor dan pendekatan lexicon (studi kasus: mobil toyota),” *Repository.Uinjkt.Ac.Id*, 2019.
- [9] S. S. Salim and J. Mayary, “ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP DOMPET ELEKTRONIK DENGAN METODE LEXICON BASED DAN K – NEAREST NEIGHBOR,” *J. Ilm. Inform. Komput.*, vol. 25, no. 1, 2020, doi: 10.35760/ik.2020.v25i1.2411.
- [10] Z. Arifin, “Penerapan Metode Knn (K-Nearest Neighbor) Dalam Sistem Pendukung Keputusan Penerimaan Kip (Kartu Indonesia Pintar) Di Desa Pandean Berbasis Web Dan Mysql,” *NJCA (Nusantara J. Comput. Its Appl.)*, vol. 4, no. 1, 2019, doi: 10.36564/njca.v4i1.101.
- [11] A. P. Giovani, A. Ardiansyah, T. Haryanti, L. Kurniawati, and W. Gata, “ANALISIS SENTIMEN APLIKASI RUANG GURU DI TWITTER MENGGUNAKAN ALGORITMA KLASIFIKASI,” *J. Teknoinfo*, vol. 14, no. 2, 2020, doi: 10.33365/jti.v14i2.679.
- [12] C. F. Suharno, M. A. Fauzi, and R. S. Perdana, “Klasifikasi Teks Bahasa Indonesia Pada Dokumen Pengaduan Sambat Online Menggunakan Metode K-Nearest Neighbors Dan Chi-square,” *Syst. Inf. Syst. Informatics J.*, vol. 3, no. 1, 2017, doi: 10.29080/systemic.v3i1.191.
- [13] R. Arief and K. Imanuel, “ANALISIS SENTIMEN TOPIK VIRAL DESA PENARI PADA MEDIA SOSIAL TWITTER DENGAN METODE LEXICON BASED,” *J. Ilm. Matrik*, vol. 21, no. 3, 2019, doi: 10.33557/jurnalmatrik.v21i3.727.
- [14] D. Muhidin and A. Wibowo, “Perbandingan Kinerja Algoritma Support Vector Machine dan K-Nearest Neighbor Terhadap Analisis Sentimen Kebijakan New Normal,” *STRING (Satuan Tulisan Ris. dan Inov. Teknol.)*, vol. 5, no. 2, 2020, doi: 10.30998/string.v5i2.6715.
- [15] R. R. A. Siregar, Z. U. Siregar, and R. Arianto, “KLASIFIKASI SENTIMENT ANALYSIS PADA KOMENTAR PESERTA DIKLAT MENGGUNAKAN METODE K-NEAREST NEIGHBOR,” *KILAT*, vol. 8, no. 1, 2019, doi: 10.33322/kilat.v8i1.421.
- [16] J. A. Septian, T. M. Fahrudin, and A. Nugroho, “Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF - IDF dan K - Nearest Neighbor,” *J. Intell. Syst. Comput.*, no. September, 2019.

- [17] N. Maleki, Y. Zeinali, and S. T. A. Niaki, "A k-NN method for lung cancer prognosis with the use of a genetic algorithm for feature selection," *Expert Syst. Appl.*, vol. 164, 2021, doi: 10.1016/j.eswa.2020.113981.
- [18] J. Xu, Y. Zhang, and D. Miao, "Three-way confusion matrix for classification: A measure driven view," *Inf. Sci. (Ny)*, vol. 507, pp. 772–794, Jan. 2020, doi: 10.1016/j.ins.2019.06.064.
- [19] "Confusion Matrix." <https://socs.binus.ac.id/2020/11/01/confusion-matrix/> (accessed Jun. 12, 2021).