

# Analisis Sentimen Masyarakat Terhadap Pembelajaran Daring di Era Pandemi Covid-19 pada Media Sosial Twitter Menggunakan Ekstraksi Fitur *Countvectorizer* dan Algoritma *K-Nearest Neighbor*

Mus Priandi<sup>1</sup>, Painem<sup>2</sup>

Teknik Informatika / Fakultas Teknologi Informasi

Jl. Cildeug Raya Petukangan Utara, Pesanggarahan, Jakarta Selatan

<sup>1</sup>1711501559@student.budiluhur.ac.id, <sup>2</sup>painem@budiluhur.ac.id

**Abstrak.** Penelitian ini bertujuan untuk melakukan analisis pandangan masyarakat terhadap sistem pembelajaran daring pada media sosial Twitter. Metode yang digunakan adalah analisis sentimen melalui pendekatan machine learning disertai fitur kamus sentimen, dengan ekstraksi fitur menggunakan *CountVectorizer* dan algoritma klasifikasi *K-Nearest Neighbor*. Dataset yang digunakan bersumber dari media sosial Twitter berupa kicauan (tweet) berbahasa Indonesia yang diperoleh melalui fitur pencarian dengan kata kunci ‘pembelajaran jarak jauh’, ‘belajar dari rumah’, ‘#belajaronline’, ‘#belajardarirumah’, ‘#belajardirumah’, dan ‘#kuliahonline’. Hasil analisis berdasarkan 3.954 tweet menunjukkan sentimen positif sebesar 76.56% dan sentimen negatif sebesar 23.44% pada periode Desember 2020, sementara hasil pengujian terbaik diperoleh menggunakan nilai  $K=7$ , dengan nilai akurasi sebesar 71%, presisi sebesar 74% dan recall sebesar 86%.

**Kata Kunci:** Analisis Sentimen, Twitter, Pembelajaran Daring, *K-Nearest Neighbor*.

## 1 Pendahuluan

Sejak 11 Maret 2020, pandemi Covid-19 telah berlangsung di seluruh dunia hingga saat ini [1]. Covid-19 menyebar dengan sangat cepat dan telah melanda 215 negara di dunia, termasuk Indonesia [2]. Penyebaran virus melalui kontak fisik memaksa semua negara untuk menerapkan pembatasan sosial, interaksi fisik dan berbagai protokol kesehatan yang ketat. Pemerintah Indonesia juga telah mengeluarkan berbagai kebijakan nasional untuk mengurangi dampak pandemi Covid-19 bagi masyarakat, mulai dari kebijakan Bekerja Dari Rumah (BDR), Sekolah Dari Rumah (SDR), Pembatasan Sosial Berskala Besar (PSBB) hingga Pemberlakuan pembatasan Kegiatan Masyarakat (PPKM) [3].

Pendidikan merupakan salah satu bidang yang terkena dampak pandemi Covid-19. Kementerian Pendidikan dan Kebudayaan Republik Indonesia melalui menternya Nadiem Makarim telah mengeluarkan kebijakan tentang pelaksanaan pendidikan dalam masa darurat Covid-19. Kebijakan tersebut menjelaskan tentang pelaksanaan proses Belajar Dari Rumah (BRD) secara daring atau online. Karenanya, seluruh institusi pendidikan diminta untuk menghentikan proses belajar mengajar baik di sekolah maupun di kampus dan menggantinya dengan sistem belajar jarak jauh. Hal ini mengakibatkan semua model pembelajaran saat ini harus berlangsung secara daring atau online dengan bantuan alat perantara seperti handphone, komputer, atau laptop [1]. Pengajar, murid, dan orang tua harus menyesuaikan diri dengan model pembelajaran tersebut. Bagi sebagian pengajar yang tidak mahir dalam penggunaan teknologi akan merasa terkejut dan harus segera beradaptasi, demikian juga murid juga orang tua. Menurut Hadion Wijoyo [4], diketahui bahwa pengajar menyenangi kelas daring sebesar 67% sedangkan yang lainnya lebih menyenangi kelas luring dikarenakan membutuhkan waktu lebih dalam mempersiapkan bahan ajar di kelas daring termasuk pemahaman perangkat IT yang digunakan. Sistem pembelajaran yang semula dianggap sebagai solusi mulai menuai beragam pendapat dari masyarakat.

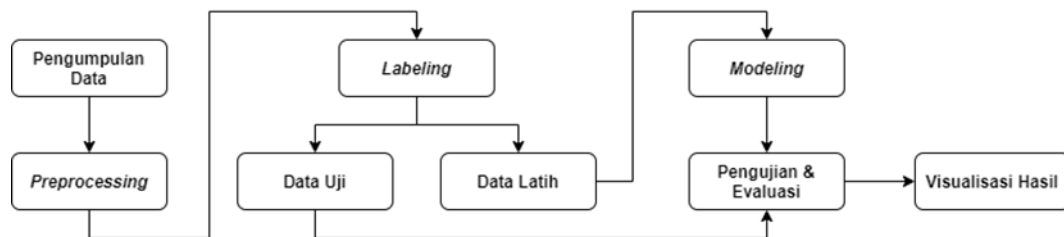
Menurut Ronal Watrionthos dalam [1] menyatakan bahwa pendapat (sentimen) masyarakat terhadap pembelajaran daring cenderung mengarah pada hasil sentimen yang negatif sebesar 83% pada bulan Juli 2020. Dalam penelitian tersebut juga dilakukan analisis emosi, menunjukkan bahwa ‘trust’ atau kepercayaan sangat mendominasi yang menandakan kepercayaan terhadap pembelajaran daring telah jauh menurun. Penelitian lain yang pernah dilakukan terkait analisis sentimen diantaranya oleh [5], menyatakan bahwa algoritma *K-Nearest Neighbor* (KNN) mampu memperoleh nilai akurasi 79,99% dengan nilai  $K=23$ . Penelitian lain yang berjudul Perbandingan Metode Naive Bayes, KNN Dan Decision Tree Terhadap Analisis Sentimen Transportasi KRL Commuter Line [6], menyatakan

bahwa KNN dapat digunakan untuk analisis sentimen dengan nilai akurasi sebesar 80% terhadap 127 data dan mampu mengimbangi algoritma Naive Bayes Classifier. KNN juga digunakan oleh Novelty dan Adiwijaya [7] dalam penelitian yang berjudul *Sentiment Analysis on Movie Reviews Using Information Gain and K-Nearest Neighbor*, untuk melakukan analisis sentimen terhadap dataset review film dengan total 2000 data, memperoleh hasil yang baik pada  $K=3$  dengan nilai akurasi sebesar 96.8%.

Penelitian ini bertujuan untuk melakukan analisis sentimen masyarakat terhadap pembelajaran daring pada periode Desember 2020. Metode yang digunakan adalah dengan melakukan analisis sentimen melalui pendekatan machine learning disertai fitur kamus sentimen, dengan ekstraksi fitur menggunakan *CountVectorizer* dan algoritma klasifikasi *K-Nearest Neighbor*. Dataset yang digunakan berupa teks kicauan (tweet) yang bersumber pada media sosial Twitter dengan kata kunci 'pembelajaran jarak jauh', 'belajar dari rumah', '#belajaronline', '#belajardarirumah', '#belajardirumah', dan '#kuliahonline'. Pengumpulan dataset dilakukan pada tanggal 1 Desember 2020 hingga 31 Desember 2020. Tercatat sebanyak 4.314 tweet yang diperoleh menggunakan kata kunci dan rentang tanggal yang diusulkan.

## 2 Metode Penelitian

Perencanaan alur analisis sentimen merupakan sebuah gambaran umum yang menunjukkan alur dari penelitian yang dilakukan. Alur penelitian tersebut dapat dilihat pada Gambar 1 di bawah ini.



**Gambar 1.** Alur Penelitian

Berdasarkan Gambar 1, pengumpulan data dilakukan melalui proses *crawling* untuk mendapatkan dataset berupa kicauan (*tweet*). Selanjutnya, *tweet* yang telah diperoleh dalam bentuk excel, kemudian dimasukkan kedalam basis data (*database*) untuk dilakukan proses *preprocessing*, pada proses *preprocessing* dilakukan penyaringan, pembuangan dan perbaikan kata. Hasil dari proses *preprocessing* menghasilkan kalimat yang lebih terstruktur (*clean text*) yang kemudian digunakan pada tahap selanjutnya. *Clean text* yang diperoleh dari proses *preprocessing* akan diproses dalam tahap *labeling* untuk menentukan kelas (label) berupa sentimen positif atau negatif, kemudian *tweet* yang telah berlabel akan dibagi menjadi dua (2) buah bagian antara lain: data uji dan data latih. Data latih merupakan data yang berfungsi sebagai pembangun pengetahuan untuk proses klasifikasi, proses pembangunan pengetahuan tersebut dilakukan melalui proses modeling dan menghasilkan sebuah model latih menggunakan data latih yang tersedia. Sementara data uji merupakan data yang disiapkan untuk menguji tingkat keakuratan model latih yang dihasilkan oleh proses modeling, proses pengujian tingkat akurasi tersebut dilakukan melalui proses pengujian dan evaluasi. Setelah dilakukan proses pengujian dan evaluasi, hasil penerapan metode akan divisualisasikan ke dalam bentuk persentase dan grafik.

### 2.1 Pengumpulan Data

Pengumpulan data dilakukan melalui proses *crawling* menggunakan pustaka *Tweepy*. *Crawling* merupakan proses mengumpulkan data dari sebuah laman dan menyimpannya untuk diatur dan dianalisis lebih lanjut [8]. Proses *crawling* tersebut memungkinkan untuk mengakses dan menambang data (*tweet*) yang bersumber dari media sosial Twitter. Periode waktu penambangan data dimulai pada tanggal 1 Desember 2020 hingga 31 Desember 2020,

sehingga diperoleh total dataset sejumlah 4.314 data (*tweet*). Kata kunci yang digunakan antara lain: ‘pembelajaran jarak jauh’, ‘belajar dari rumah’, ‘#belajaronline’, ‘#belajardarirumah’, ‘#belajardirumah’, dan ‘#kuliahonline’.

Adapun atribut data yang dikumpulkan melalui proses *crawling* antara lain:

- 2.1.1. *Tweet id*, yaitu atribut *unique* dari setiap *tweet*, digunakan sebagai pembeda (*primary key*) antara *tweet* satu dengan yang lainnya.
- 2.1.2. *Full text*, yaitu atribut terpenting dalam proses pengumpulan data, yang berisikan teks opini/pendapat/kicauan dari masyarakat dengan jumlah panjang maksimal sebanyak 280 karakter [5] [11].
- 2.1.3. *Created at*, berisikan format waktu yyyy-mm-dd hh:mm:ss, yang merupakan waktu dibuatnya sebuah *tweet*.
- 2.1.4. *User screen name*, yaitu *username* dari akun yang mempublikasikan sebuah *tweet*.

## 2.2 Preprocessing

Data *tweet* yang diperoleh berasal dari media sosial Twitter belum bisa dipakai karena disebabkan masih mempunyai bentuk teks atau kalimat yang tidak beraturan (tidak terstruktur). Selain itu, data masih memiliki kotor dan memiliki banyak kata yang tidak diperlukan. Oleh karena itu, perlu dilakukan proses *preprocessing* terlebih dahulu. *Preprocessing* merupakan bagian dari proses penggalian teks yang dilakukan untuk menghapus kata atau kalimat yang tidak diperlukan dari suatu kumpulan teks. Selain itu, proses ini bertujuan untuk menghindari data yang kurang sempurna, gangguan pada data, dan data yang tidak konsisten [9].

Adapun tahapan *preprocessing* yang dilakukan adalah sebagai berikut:

- 2.2.1. *Case folding*, yaitu proses pengubahan seluruh huruf menjadi huruf kecil (*lowercase*) [10] [11].
- 2.2.2. *Cleansing*, yaitu proses yang bertujuan untuk menghapus atribut yang tidak diperlukan untuk analisis analisis [1] [10] [11].
- 2.2.3. Mengubah *slang word*, yaitu proses pengubahan kata yang tidak sesuai dengan ejaan bahasa Indonesia yang baku (EYD) baik berupa kata singkatan, kata gaul atau modern, ataupun kesalahan salah eja [12].
- 2.2.4. Menghapus *stop word*, yaitu proses penghapusan kata yang tidak berpengaruh atau kurang bermakna namun sering ditemui dalam dokumen atau kalimat [1] [10].
- 2.2.5. *Stemming*, yaitu proses mengubah kata berimbuhan menjadi kata dasar [1] [11].

## 2.3 Labeling

*Labeling* atau pelabelan merupakan proses pemberian kelas berdasarkan ciri atau karakteristik yang terkandung dalam sebuah dokumen atau kalimat. Dalam penelitian ini proses pelabelan akan memberikan kelas pada tiap *tweet* dengan label positif atau label negatif. Label positif menunjukkan bahwa *tweet* yang bersangkutan merupakan *tweet* yang cenderung bersifat setuju/mendukung/menerima proses berlangsungnya pembelajaran jarak jauh, sementara label negatif menunjukkan bahwa *tweet* tersebut cenderung bersifat menyangkal/menolak/menampik proses berlangsungnya pembelajaran jarak jauh.

Adapun cara menentukan kelas pada *tweet* dapat dilakukan dengan dua (2) cara antara lain: pelabelan manual dengan melabeli kalimat berdasarkan subjektivitas orang banyak dan pelabelan dengan pendekatan kamus sentimen. Tahapan melakukan pelabelan dengan pendekatan kamus sentimen antara lain perhitungan skor sentimen dan pemberian kelas sentimen.

Adapun cara utama yang digunakan dalam penelitian ini untuk memberikan label sentiment yaitu dengan menggunakan pendekatan kamus sentimen, dengan langkah-langkah sebagai berikut:

### 2.3.1 Perhitungan skor sentimen

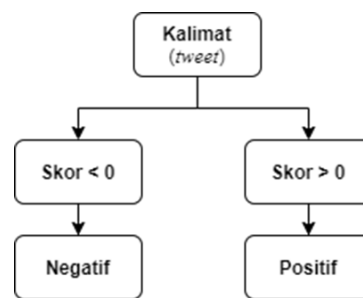
Untuk melakukan perhitungan skor sentimen digunakan pendekatan kamus sentimen. Kamus tersebut berisikan kata sentimen positif dan negatif. Skor suatu kata akan bernilai +1 jika kata tersebut adalah kata opini positif, dan

bernilai -1 jika kata tersebut adalah kata opini negatif [13] [14]. Proses perhitungan nilai skor sentimen diperoleh menggunakan Persamaan (1).

$$\text{Skor} = (\sum(\text{kata positif}) - (\sum\text{kata negatif})) \quad (1)$$

### 2.3.2 Pemberian kelas sentimen

Proses pelabelan kelas sentimen dilakukan secara otomatis. Proses pelabelan dilakukan dengan pendekatan kamus sentimen yang memberikan kelas pada kalimat (*tweet*) berdasarkan skor sentimen positif atau negatif. Jika kalimat mempunyai skor  $> 0$  akan masuk ke dalam kelas positif. Sebaliknya, jika kalimat mempunyai skor  $< 0$  akan masuk ke dalam kelas negatif [10] [13] [8]. Selanjutnya jika kalimat mempunyai skor  $= 0$  maka akan diabaikan sehingga penentuan kelas sentimen dilakukan secara manual. Adapun proses pemberian kelas dapat dilihat pada Gambar 2.



**Gambar 2.** Pemberian kelas sentiment

Cara pemberian kelas secara manual dilakukan kepada *tweet-tweet* yang memiliki nilai skor  $= 0$ , cara tersebut dilakukan dengan meminta bantuan kepada beberapa pakar yang merupakan pelajar dan mahasiswa. Pakar melakukan pelabelan *tweet* satu per-satu berdasarkan objektivitas masing-masing. Pakar diberikan daftar *tweet* yang siap untuk dilabeli dengan cara memilih satu dari tiga *label* (positif, negatif, netral) pada tiap-tiap *tweet*. Kemudian hasil pelabelan dari tiap-tiap orang akan digunakan untuk mencari kelas sentimen yang dominan muncul/dipilih, jika dominannya adalah kelas positif maka *tweet* tersebut akan dikategorikan kedalam kelas positif; jika dominannya adalah kelas negatif maka *tweet* tersebut akan dikategorikan kedalam kelas negatif; jika dominannya bukan kelas positif atau bukan kelas negatif atau jumlah kelas positif dan negatifnya sama maka *tweet* tersebut tidak diberikan label dan tidak digunakan dalam proses analisis sentimen.

### 2.4 Pembagian Data

Data yang telah memiliki label atau berlabel akan dibagi menjadi dua (2) buah bagian antara lain: data uji dan data latih. Proses pembagian data dilakukan dengan membagi dataset menjadi 90% data latih dan 10% menjadi data uji.

### 2.5 Modeling

*Modeling* merupakan proses pembuatan pengetahuan berdasarkan data latih yang telah tersedia. Data latih yang dijadikan model dipilih dengan teknik sampling kuota (*quota sampling*). Quota Sampling merupakan teknik sampling yang menentukan jumlah sampel dari populasi yang memiliki ciri atau kriteria tertentu hingga jumlah kuota yang diinginkan tercapai [9].

Proses *modelling* dilakukan menggunakan ekstraksi fitur *CountVectorizer*, digunakan untuk menghitung frekuensi kata dalam dokumen atau kalimat kemudian direpresentasikan ke dalam bentuk vektor. Bentuk representasi vektor tersebut kemudian disimpan sebagai model latih dalam bentuk file JSON (.json).

## 2.6 Pengujian dan Evaluasi

Pengujian dilakukan menggunakan K-Nearest Neighbor (KNN) sebagai algoritma klasifikasi *supervised learning* berbasis jarak. Algoritma ini bekerja dengan cara membandingkan jarak antara data uji dengan semua data latih yang ada [6] [5] [7]. Untuk menghitung jarak antara data digunakan perhitungan euclidean distance dengan rumus sebagai berikut:

$$d_{(x,y)} = \sqrt{\sum_{i=1}^n (X_i - Y_i)^2} \quad (2)$$

Pengujian dilakukan dengan memperbandingkan hasil prediksi yang dilakukan oleh sistem (aplikasi) dengan hasil pelabelan yang dilakukan oleh pakar. Data prediksi berupa hasil klasifikasi yang dihasilkan oleh aplikasi yang dibangun, sedangkan data aktual berupa yang didapatkan melalui proses pelabelan [15]. Dalam penelitian ini, pengujian dilakukan pada sebuah model terhadap data uji yang tersedia dengan variasi nilai K, yaitu: K=3, K=5, K=7, K=9, dan K=11. Hasil dari pengujian tersebut akan dievaluasi menggunakan confusion matrix untuk mengukur tingkat akurasi, presisi dan recall.

## 3 Hasil dan Pembahasan

### 3.1. Dataset

*Dataset* penelitian bersumber dari media sosial Twitter berupa data teks kicauan (*tweet*). *Dataset* tersebut diperoleh secara rutin melalui proses *crawling* menggunakan pustaka *Tweepy*, dimulai pada tanggal 1 Desember 2020 hingga 31 Desember 2020, sehingga diperoleh total *dataset* sejumlah 4.314 data (*tweet*). Kata kunci yang digunakan antara lain: ‘pembelajaran jarak jauh’, ‘belajar dari rumah’, ‘#belajaronline’, ‘#belajardarirumah’, ‘#belajardirumah’, dan ‘#kuliahonline’. Adapun atribut yang dimiliki oleh setiap *dataset* antara lain: *tweet id*, *full text (tweet)*, *created at*, *user screen name (username)*.

### 3.2. Preprocessing Data

Seperti sudah dijelaskan pada bagian sebelumnya, data asli yang diperoleh dari media sosial Twitter dilakukan pengolahan terlebih dahulu (*preprocessing*). Adapun langkah-langkah *preprocessing* yang dilakukan pada penelitian ini terdiri dari tahap *case folding*, *cleansing*, mengubah *slang word*, menghapus *stop word*, *stemming*. Pada Tabel 1 di bawah ini menunjukkan proses dari tahap *preprocessing* data.

Tabel 1. Proses *preprocessing* data

| Tahap               | Input  | Output   |
|---------------------|--|--|
| <i>Case folding</i> | Jadwal Program "BELAJAR DARI RUMAH" Hari Kamis, 31 Desember 2020 pukul 08.00-11.00 WIB<br><br>#BelajarDariRumah<br>#MediaPemersatuBangsa<br>#TVRI @TVRINasional<br><a href="https://t.co/Jg7L7Kw9Bv">https://t.co/Jg7L7Kw9Bv</a> | jadwal program "belajar dari rumah" hari kamis, 31 desember 2020 pukul 08.00-11.00 wib<br><br>#belajardarirumah<br>#mediapemersatubangsa<br>#tvri @tvrinasional<br><a href="https://t.co/jg717kw9bv">https://t.co/jg717kw9bv</a> |
| <i>Cleansing</i>    | jadwal program "belajar dari rumah" hari kamis, 31 desember 2020 pukul 08.00-11.00 wib<br><br>#belajardarirumah  | jadwal program belajar dari rumah hari kamis desember pukul wib  |

|                               |  |  |
|-------------------------------|--|--|
|                               | #mediapemersatubangsa<br>#tvri @tvrinasional <a href="https://t.co/jg717kw9bv">https://t.co/jg717kw9bv</a> |  |
| Mengubah<br><i>slang word</i> | jadwal program belajar dari rumah hari<br>kamis desember pukul wib   | jadwal program belajar dari rumah hari<br>kamis desember pukul wib |
| Menghapus<br><i>Stop word</i> | jadwal program belajar dari rumah hari<br>kamis desember pukul wib   | jadwal program belajar rumah kamis<br>desember pukul               |
| <i>Stemming</i>               | jadwal program belajar rumah kamis<br>desember pukul   | jadwal program ajar rumah kamis<br>desember pukul                  |

### 3.3. Labeling Data

Tahap *labeling* data dikelompokkan menjadi dua kategori atau 2 kelas sentimen yaitu sentimen positif dan negatif. Pelabelan dilakukan secara otomatis dengan melakukan perhitungan nilai tertentu. Suatu dokumen dapat masuk ke dalam kelas positif atau negatif dengan memanfaatkan kumpulan kata berbahasa Indonesia atau sebuah kamus yang terdiri dari kumpulan kata positif dan kumpulan kata negatif.

Berdasarkan *tweet* “semangat teman teman ikut aktif bantu dalam proses ajar jarak jauh”, terdapat 8 kata positif, dan 1 kata negatif yang terdeteksi oleh kamus sentimen. Ilustrasinya perolehan nilainya dapat dilihat pada Tabel 2 sebagai berikut.

**Tabel 2.** Perhitungan skor sentimen

| <i>Tweet</i>  | Kata Positif  | Kata Negatif |
|---|---|--------------|
| semangat teman teman ikut aktif<br>bantu dalam proses ajar jarak jauh | semangat (1)<br>teman (2)<br>ikut (1)<br>aktif (1)<br>bantu (1)<br>proses (1)<br>ajar (1) | jauh (1)     |
| <b>Jumlah</b>   | 8   | 1            |

Berdasarkan tabel diperoleh hasil perhitungan : skor = (jumlah kata positif) – (jumlah kata negatif) , skor = 8 – 2 maka skor = 6. Berdasarkan Gambar 2, maka *tweet* “semangat teman teman ikut aktif bantu dalam proses ajar jarak jauh” yang memiliki nilai skor = 6 akan dikategorikan ke dalam kelas positif, karena nilai skor > 0.

### 3.4. Pengujian dan Evaluasi

Pada tahap pengujian dan evaluasi dalam penelitian yang dilakukan ini adalah dengan menggunakan model latih serta data uji. Berdasarkan oleh model latih yang terdiri atas 1500 data latih berlabel dengan ciri: 750 *tweet* latih positif dan 750 *tweet* latih negatif dan data uji sebanyak 395 data, berikut pada Tabel 3 merupakan hasil sampel prediksi oleh algoritma KNN dengan nilai K=3.

**Tabel 3.** Sampel data hasil prediksi

| No | <i>Tweet</i> | Label aktual | Label prediksi |
|----|--------------|--------------|----------------|
|----|--------------|--------------|----------------|

|     |  |         |         |
|-----|--|---------|---------|
| 1   | Selain untuk memperkuat kompetensi literasi dan numerasi, tujuan lain program BDR adalah untuk membangun ikatan emosional dalam keluarga. <a href="https://t.co/Hks8kSVh13">https://t.co/Hks8kSVh13</a>  | positif | positif |
| 2   | @schfess Kontra, udah capek batin belajar dari rumah, materi gak ada yang masuk dan karna gua SMK di mana harus banyak praktek.  | negatif | negatif |
| 3   | Pandemi juga memaksa adik-adik kita belajar online dari rumah. Sayangnya, tidak semua dari mereka memiliki fasilitas internet dan gawai yang mendukung. Untungnya, banyak orang baik di luar sana yang tergerak membantu adik-adik agar bisa tetap bisa belajar  | positif | negatif |
| ... | ...  | ...     | ...     |
| 395 | Seri Kepribadian Muslim 3 : Bersih dan Sehat - Diskon 20% menjadi Rp.20000 Gratis ongkos kirim hingga 40rb keseluruhan Indonesia. #mainanbukuanak #promo #dirumahaja #belajardirumah #buku #membaca #booklover #bookstorm<br>SINOPSIS: Aku suka bersih. Aku suka | positif | positif |

Pada Tabel 3 Sampel data hasil prediksi, kolom *label* aktual merupakan data *label* yang diperoleh melalui proses *labeling*, sementara *label* prediksi merupakan data *label* hasil dari proses klasifikasi menggunakan KNN. Keseluruhan hasil prediksi data uji (395 data *tweet*) kemudian direpresentasikan ke dalam *confusion matrix*. Representasi *confusion matrix* untuk K=3 yang terbentuk dapat terlihat pada Tabel 4 berikut.

**Tabel 4.** *Confusion matrix* pengujian K=3

|                |         | Nilai Aktual |         |
|----------------|---------|--------------|---------|
|                |         | positif      | negatif |
| Nilai Prediksi | positif | 211          | 70      |
|                | negatif | 50           | 64      |

Berdasarkan Tabel 4 *Confusion matrix* pengujian K=3, maka perolehan nilai akurasi, presisi dan *recall* dapat dilihat pada Tabel 5 berikut.

**Tabel 5.** Nilai pengujian K=3

| Pengujian     |                                 |             |
|---------------|---------------------------------|-------------|
| Akurasi       | $= \frac{211+64}{211+64+70+50}$ | 0.7 (70 %)  |
| Presisi       | $= \frac{211}{211+70}$          | 0.75 (75 %) |
| <i>Recall</i> | $= \frac{211}{211+50}$          | 0.81 (81 %) |



Pengujian di atas dilakukan secara berulang dengan variasi nilai K yang berbeda-beda. Sehingga dapat diketahui hasil pengujian secara keseluruhan adalah seperti Tabel 6 berikut.

**Tabel 6.** Hasil pengujian dan evaluasi

|               | K=3  | K=5  | K=7  | K=9  | K=11 |
|---------------|------|------|------|------|------|
| Akurasi       | 0.7  | 0.7  | 0.71 | 0.68 | 0.67 |
| Presisi       | 0.75 | 0.75 | 0.74 | 0.72 | 0.72 |
| <i>Recall</i> | 0.81 | 0.84 | 0.86 | 0.84 | 0.81 |

Berdasarkan Tabel 6 di atas diketahui bahwa nilai akurasi terendah diperoleh dengan nilai K=11, yaitu akurasi 0.67 (67%), presisi 0.72 (72%) dan *recall* 0.81 (81%), sedangkan nilai akurasi tertinggi didapatkan dengan nilai K=7, yaitu akurasi 0.71 (71%), presisi 0.74 (74%) dan *recall* 0.86 (86%). Dalam penelitian ini, nilai optimal dari K adalah 7 (71%), sedangkan dalam penelitian lain yaitu: penelitian [7] yang berjudul *Sentiment Analysis on Movie Reviews Using Information Gain and K-Nearest Neighbor*, nilai optimal dari K adalah 3 (akurasi 96.8%) dan penelitian [5] yang berjudul Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-Nearest Neighbor, nilai optimal yang diperoleh adalah K=23 (akurasi 79,99%). Dari hasil-hasil penelitian tersebut disimpulkan bahwa nilai K bersifat tidak tentu, harus dicari dan bergantung pada jenis, jumlah dan bobot dari *dataset* yang digunakan.

## 4 Kesimpulan

Berdasarkan *dataset* yang diproses, arah pandangan (sentimen) masyarakat Indonesia terhadap pembelajaran daring cenderung ke arah sentimen yang baik (positif) sebesar 76.56% pada periode tanggal 1 Desember 2020 sampai dengan 31 Desember 2020. *Dataset* yang digunakan sebanyak 3.954 *tweet*, di mana 3.027 data memiliki sentimen positif dan 927 data sisanya bersentimen negatif.

Penggunaan ekstraksi fitur *CountVectorizer* dan algoritma *K-Nearest Neighbor* (KNN) dalam melakukan analisis sentimen dapat berjalan dengan baik, dengan nilai pengujian dan evaluasi tertinggi yang diperoleh sebesar: akurasi 71%, presisi 74% dan *recall* 86% menggunakan nilai K=7.

## Referensi

- [1] R. Watrionthos, "Analisis Pembelajaran Daring di Era Pandemic Covid-19," *Merdeka Kreat. di Era Pandemi Covid-19*, pp. 55–64, 2020.
- [2] A. Sadikin and A. Hamidah, "Pembelajaran Daring di Tengah Wabah Covid-19 (Online Learning in the Middle of the Covid-19 Pandemic)," *BIODIK J. Ilm. Pendidik. Biol.*, vol. 6, no. 1, pp. 214–224, 2020, doi: <https://doi.org/10.22437/bio.v6i2.9759>.
- [3] A. Ristyawati, "Efektifitas Kebijakan Pembatasan Sosial Berskala Besar Dalam Masa Pandemi Corona Virus 2019 oleh Pemerintah Sesuai Amanat UUD NRI Tahun 1945," *Adm. Law Gov. J.*, vol. 3, no. 2, pp. 240–249, 2020.
- [4] H. Wijoyo, "Guru Milenial dan Covid-19," *Merdeka Kreat. di Era Pandemi Covid-19*, pp. 27–41, 2020.
- [5] 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.*, vol. 1, no. 1, pp. 43–49, 2019.
- [6] N. T. Romadloni, I. Santoso, and S. Budilaksono, "Perbandingan Metode Naive Bayes, KNN Dan Decision Tree Terhadap Analisis Sentimen Transportasi KRL Commuter Line," *J. IKRA-ITH Inform.*, vol. 3, no. 2, pp. 1–9, 2019.
- [7] N. O. F. Daeli and Adiwijaya, "Sentiment Analysis on Movie Reviews Using Information Gain and K-Nearest Neighbor," *J. Data Sci. Its Appl.*, vol. 3, no. 1, pp. 1–7, 2020, doi: 10.34818/JDSA.2020.3.22.
- [8] F. Nurulbaiti and M. S. Retno Subekti, "Analisis Sentimen Terhadap Data Tweet Untuk Badan Penyelenggara Jaminan Sosial (BPJS) Menggunakan Program R," *J. Pendidik. Mat. dan Sains*, pp. 1–9, 2020.



- [9] F. V. Sari and A. Wibowo, "Analisis Sentimen Pelanggan Toko Online Jd.Id Menggunakan Metode Naïve Bayes Classifier Berbasis Konversi Ikon Emosi," *J. SIMETRIS*, vol. 10, no. 2, pp. 681–686, 2019.
- [10] E. B. Santoso and A. Nugroho, "Analisis Sentimen Calon Presiden Indonesia 2019 Berdasarkan Komentar Publik di Facebook," *J. Eksplora Inform.*, vol. 9, no. 1, pp. 60–69, 2019, doi: 10.30864/eksplora.v9i1.254.
- [11] S. N. J. Fitriyyah, N. Safriadi, and E. E. Pratama, "Analisis Sentimen Calon Presiden Indonesia 2019 dari Media Sosial Twitter Menggunakan Metode Naive Bayes," *JEPIN (Jurnal Edukasi dan Penelit. Inform.*, vol. 5, no. 3, pp. 279–285, 2019.
- [12] P. Antinasari, R. S. Perdana, and M. A. Fauzi, "Analisis Sentimen Tentang Opini Film Pada Dokumen Twitter Berbahasa Indonesia Menggunakan Naive Bayes Dengan Perbaikan Kata Tidak Baku," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 1, no. 12, pp. 1733–1741, 2017.
- [13] G. A. Buntoro, "Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter," *Integer J.*, vol. 2, no. 1, pp. 32–41, 2017.
- [14] B. Liu, M. Hu, and J. Cheng, "Opinion Observer : Analyzing and Comparing Opinions on the Web," *Proc. 14th Int. World Wide Web Conf.*, 2005.
- [15] D. H. Wahid and A. SN, "Peringkasan Sentimen Esktraktif di Twitter Menggunakan Hybrid TF-IDF dan Cosine Similarity," *J. IJCCS*, vol. 10, no. 2, pp. 207–218, 2017.