

Penerapan Algoritma *K-Nearest Neighbor* pada *Twitter* untuk Analisis Sentimen Masyarakat Terhadap Larangan Mudik 2021

Diah Ayu Lestari¹, Deni Mahdiana^{2*}

^{1,2}Program Studi Sistem Informasi, Fakultas Teknologi Informasi, Universitas Budi Luhur, Jakarta, Indonesia

Email: ¹diahayu0899@gmail.com, ^{2*}deni.mahdiana@budiluhur.ac.id

(* : coresponding author)

Abstrak. Media sosial *Twitter* merupakan media yang banyak digunakan oleh masyarakat dalam menyampaikan sebuah opini yang sedang hangat dibahas. Kebijakan larangan mudik yang diterapkan oleh pemerintah saat ini belum diketahui opini atau pendapat masyarakat terhadap pelaksanaan larangan mudik 2021 ini, sehingga pemerintah kesulitan dalam mengevaluasi kebijakan larangan mudik tersebut. Penelitian ini akan melakukan analisis sentimen menggunakan algoritma *K-Nearest Neighbor* (K-NN) dan menerapkan metodologi *Cross-Industry Standard Process for Data Mining* (CRIPS-DM). Pada penelitian ini data bersih yang digunakan berjumlah 4.799 *tweet*, yang diambil dari media sosial *twitter* pada 04 April 2021 – 17 Mei 2021 dengan sentimen positif berjumlah 834 *tweet* dan 3.965 *tweet* sentimen negatif. Penelitian ini menghasilkan bahwa K-NN dapat diimplementasikan dengan baik dikarenakan mencapai nilai akurasi sebesar 86.67 % dengan nilai *recall* 39.52 %, *precision* 70.97 % dan *spencificity* sebesar 96.60% menggunakan *split data* perbandingan 80 untuk data *training* dan 20 untuk data *testing* dengan nilai $k=3$. Sehingga dapat dikatakan bahwa algoritma K-NN dapat mengklasifikasikan data secara benar dan baik.

Kata Kunci: Analisis Sentimen, CRIPS-DM, *K-Nearest Neighbor*, Larangan Mudik

1. Pendahuluan

Berdasarkan Kamus Besar Bahasa Indonesia (KBBI), pengertian mudik adalah pulang ke kampung halaman. Pemerintah kembali menerapkan kebijakan larangan mudik lebaran ditahun 2021 ini. Penerapan kebijakan ini diambil setelah mempertimbangkan tingginya angka penularan dan kematian *COVID-19* pasca libur panjang terutama pada perayaan Idul Fitri tahun lalu mengalami peningkatan yang signifikan [1]. Berdasarkan data Satgas *COVID-19*, hari raya Idul Fitri tahun lalu menghasilkan peningkatan rata-rata harian kasus sebanyak 68-93% atau peningkatan 413-559 kasus per hari, dengan jumlah kasus mingguan mulai dari 2.889-3.917. Sedangkan *persentase* kematian per minggu berkisar antara 28-66% atau sebanyak 6-143 kematian [2]. Untuk tanggal 08 April 2021 jumlah pasien terkonfirmasi positif yang tercatat sejak kasus pertama hingga hari ini, mencapai 1.547.376 kasus [3]. Staf Menteri Perhubungan Aditya Irawan mengemukakan alasan lain terkait pelarangan mudik, yaitu pergerakan masyarakat yang banyak dapat meningkatkan penularan *COVID-19* secara meningkat juga [4]. Oleh karena itu penerapan larangan mudik ini dimaksud agar warga tidak melakukan mobilitas secara besar-besaran untuk menghindari terjadi penambahan kasus baru *COVID-19*.

Banyak masyarakat Indonesia yang menuangkan opini di *Twitter* mengenai larangan mudik. Opini tersebut belum dapat dibedakan bersentimen negatif ataupun positif, sehingga pemerintah kesulitan dalam melihat penilaian masyarakat. Karena penelitian ini berfokus pada opini dari *Twitter*, penelitian ini menggunakan teknik *text mining* untuk memproses data yang diambil dari *Twitter*. Analisis sentimen adalah bagian dari ilmu *text mining* yang memiliki tujuan untuk menganalisis dan mengekstrak data tekstual yang berupa pendapat, evaluasi, sikap, emosi, penilaian, dan sentimen seseorang terhadap suatu barang, orang, organisasi dan masalah [5].

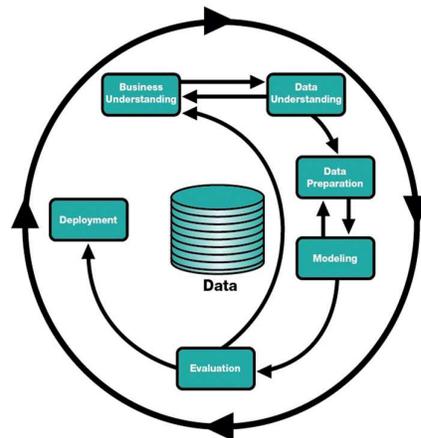
Terdapat beberapa penelitian yang telah membahas tentang sentimen analisis larangan mudik menggunakan metode Naïve Bayes dilakukan oleh [6] dan *Support Vector Machine* (SVM) dilakukan oleh [7]. Penelitian ini fokus pada penerapan metode K-NN untuk analisis sentimen larangan mudik 2021. Algoritma K-NN ini memiliki keunggulan yaitu memiliki prinsip sederhana, bekerja berdasarkan jarak terpendek dari sampel uji dan sampel latih [8]. Menurut Andi [9] K-NN juga memiliki kelebihan kepada *training dataset* yang memiliki data *training* tinggi/besar dan untuk kelemahan dari K-NN yaitu masih perlu penentuan nilai k.

Penelitian ini bertujuan untuk mengklasifikasikan terhadap *tweet* yang mengandung nilai positif dan negatif mengenai larangan mudik 2021 dengan menggunakan algoritma *K-Nearest Neighbor* (K-NN). Pada penelitian ini

hanya menggunakan sentimen positif dan negatif dikarenakan untuk melihat penilaian masyarakat baik atau buruk terhadap kebijakan larangan mudik. Hasil penelitian tersebut akan menjadi gambaran mengenai larangan mudik 2021 apakah opini publik lebih cenderung ke opini positif atau negatif dan melihat hasil akurasi yang dihasilkan dari algoritma yang digunakan yaitu *K-Nearest Neighbor* (K-NN). Sehingga nantinya hasil dari penelitian ini dapat digunakan sebagai bahan evaluasi larangan mudik 2021 oleh pemerintah untuk memperbaiki kebijakan selanjutnya dan sebagai referensi untuk penelitian selanjutnya.

2. Metode Penelitian

2.1 Metode CRIPS-DM



Gambar. 6 Proses CRIPS-DM [10]

Berdasarkan Gambar 1 dalam penelitian ini peneliti menggunakan *text mining* dengan tahapan CRISP-DM agar memenuhi standar proses yang ada dan dapat diterapkan ke dalam strategi pemecahan suatu masalah. Berikut adalah penjelasan mengenai tahapan-tahapan yang digunakan dalam penelitian ini:

a. Business Understanding

Pada tahapan ini berfokus pada pemahaman mengenai tujuan dari proyek dan kebutuhan secara persepektif bisnis, kemudian mengubah pengetahuan ini menjadi definisi masalah penambangan data dan rencana awal yang dirancang untuk mencapai tujuan. Untuk tahapan ini langkah pertama peneliti akan memahami permasalahan yang diangkat oleh penelitian ini dengan mempelajari artikel atau berita dan jurnal terkait kebijakan larangan mudik 2021 yang berlaku di Indonesia, kemudian mencari *tweet* atau *trending topic* yang berkaitan dengan larangan mudik di *twitter*.

b. Data Understanding

Pada tahapan ini dilakukan pengumpulan, pengidentifikasi, dan pemahaman data yang akan digunakan pada penelitian ini. Dataset yang digunakan dalam penelitian ini didapat dari hasil *crawling* di *twitter* menggunakan *rapidminer*. Untuk pengumpulan datanya menggunakan *keyword* yang sesuai dengan topik larangan mudik yaitu pada penelitian ini menggunakan *keyword* “larangan mudik”, “#patuhtidakmudik” dan “mudik”. Peneliti melakukan pengambilan data *tweet* dari media sosial *twitter* pada tanggal 04 April 2021 hingga 17 Mei 2021, dengan mengambil *tweet* setiap minggunya, dengan data *tweet* yang dikumpulkan sebanyak 12.502 *tweet*. Sebelum melakukan proses *crawling tweet*, peneliti membuat koneksi di *rapidminer* agar terhubung dengan *twitter*. Data yang diambil hanya data yang sesuai dengan dengan *keyword* yang sudah ditentukan.

c. Data Preparation

Pada tahapan ini data akan dibersihkan melalui *text preprococessing* yang terdiri dari beberapa tahapan yaitu[11] :

1. *Cleansing* merupakan tahapan menghapus *duplikat* dan menghapus simbol, karakter, @username, RT, dan URL yang tidak digunakan dalam data.

2. *Case folding* merupakan tahapan untuk mengubah huruf pada dataset menjadi *lower* atau huruf kecil semua.
3. *Tokenizing* merupakan tahapan untuk memisahkan kalimat menjadi perkata di *dataset*
4. *Stopwords* merupakan tahapan untuk membuang kata yang tidak diperlukan seperti kata penghubung "di", "yang", "sedang", "dan". Untuk penelitian ini peneliti membuat *stopword list* dari. Pada penelitian ini menggunakan *stopword* dari [12].
5. *Stemming* merupakan tahapan untuk menghilangkan kata imbuhan pada kalimat sehingga hanya ada kata dasar di *dataset*. Pada proses *stemming* ini peneliti akan membuat *file stem* yang berisikan kata-kata yang masih berhamburan dan mengganti dengan kata dasarnya.
6. *TF-IDF* merupakan proses untuk pembobotan kata.

Selanjutnya data akan dilakukan pelebelan dengan menggunakan kamus *lexicon*. Pelebelan *dataset* ini akan dibantu dengan *Rstudio* dengan menerapkan rumus perhitungan sentiment yang berdasarkan dari list kata yang termasuk kedalam kata positif dan negatif yang sudah ditentukan [13]. Setelah pelebelan sentimen dilakukan, maka data dilakukan proses *preprocessing* data sehingga data menjadi bersih dan dapat digunakan.

d. *Modelling*

Pada tahapan *modelling* ini peneliti akan melakukan pemodelan terhadap *dataset* yang sudah dilakukan *preprocessing*. Kemudian *dataset* tersebut akan diuji dengan pemodelan *split* dan *k-fold cross validation* untuk menghasilkan akurasi

e. *Evaluation*

Pada tahap ini peneliti akan melakukan evaluasi metode klasifikasi dengan mengukur *performance* menggunakan *confusion matrix* terhadap algoritma K-NN.

3. Hasil Dan Pembahasan

3.1 Business Understanding

Untuk tahapan *business understanding* ini peneliti akan memahami permasalahan yang diangkat selama penelitian ini dengan membaca jurnal, artikel dan berita terkait. Untuk permasalahan yang terdapat dalam penelitian ini belum diketahui opini masyarakat lebih cenderung sentimen negatif atau positif mengenai kebijakan larangan mudik 2021 yang ditetapkan oleh pemerintah. Penelitian ini akan menggunakan data yang ditarik melalui media sosial *twitter*. Kata yang sedang *trending topic* dapat dijadikan sebagai *keywords* pada saat proses *crawling data*, karena dapat memudahkan peneliti untuk mendapatkan data *tweet* sesuai dengan pembahasan yang sedang diteliti.

3.2 Data Understanding

Pada tahapan ini peneliti akan menentukan *keywords* atau kata kunci yang cocok dan berkaitan dengan topik untuk digunakan dalam proses *crawling data tweet*. Peneliti melakukan penarikan data *tweet* yang berkaitan dengan topik larangan mudik 2021 dimulai pada tanggal 04 April 2021 hingga 17 Mei 2021, dengan menggunakan beberapa *trending topic* seperti Mudik dan #PatuhTidakMudik karena masih ada hubungannya dengan pelaksanaan larangan mudik. Peneliti melakukan *crawling data* yang akan digunakan selama penelitian menggunakan media sosial *twitter* dengan menggunakan *RapidMiner* seperti yang disajikan pada Tabel 1.

Tabel 7. Jumlah Data Hasil *Crawling*

<i>Keywords</i>	<i>Tanggal</i>	<i>Jumlah</i>
Larangan Mudik	04 April 2021 – 17 Mei 2021	8.220
Larangan Mudik + Mudik	04 April 2021 – 09 Mei 2021	4.082
Larangan Mudik+ #PatuhTidakMudik	06 Mei 2021 – 9 Mei 2021	200
Jumlah		12.502

3.3 Data Preparation

Pada tahapan *data preparation* ini peneliti akan melakukan penggabungan data dari hasil *crawling* lalu melakukan *cleansing* pada data sehingga data menjadi lebih bersih. Selanjutnya akan dilakukan penghitungan skor sentiment untuk menentukan *label* kelas menggunakan kamus *lexicon* dibantu dengan *tools Rstudio*, lalu data yang sudah diberi *label* akan diproses dengan *text preprocessing*.

a. Proses *Cleansing*

Pada tahapan ini peneliti akan melakukan proses *Cleansing* dengan menggunakan *rapidminer* sebelum melakukan pemberian label pada sentimen, agar *dataset* pada file *excel* terbaru terhindar dari duplikasi data dan tanda yang tidak diperlukan. Pada proses ini akan dilakukan pembersihan dari berbagai *noise* seperti menghilangkan *link URL*, *username*, *retweet*, digit angka dan karakter. Setelah selesai menjalankan proses *Cleansing* dengan menggunakan *rapidminer*. Awalnya terdapat 12.502 data *tweet* pada file *excel*, kemudian setelah melewati proses *Cleansing* maka data *tweet* berkurang menjadi 7.641 data.

b. Pemberian Label Sentimen

Kemudian tahap selanjutnya adalah melakukan klasifikasi atau penentuan sentimen berdasarkan masing-masing *tweet*. Selama proses penentuan sentimen berlangsung, disini peneliti menggunakan kamus *lexicon* sebagai pemberian sentimen dan dibantu oleh *tools Rstudio*. Pada pemberian *labelling* ini kamus *lexicon* sudah cukup baik dalam menilai sentimen, hal ini dikarenakan kata-kata dalam kamus *lexicon* sudah cukup sesuai antara kata bersifat negatif dan positif. Berikut ini adalah contoh pemberian *label* sentimen dari sebuah *tweet*, seperti yang ditunjukkan pada Tabel 2.

Tabel 8. Perhitungan Skor Sentimen

Text	Jumlah Positif	Jumlah Negatif	Score	Label
Warga Ramairamai Protes dan Keluhkan Larangan Mudik oleh pemerintah	1	3	-2	Negatif

Maka, diperoleh hasil dari perhitungan pada Table 2:

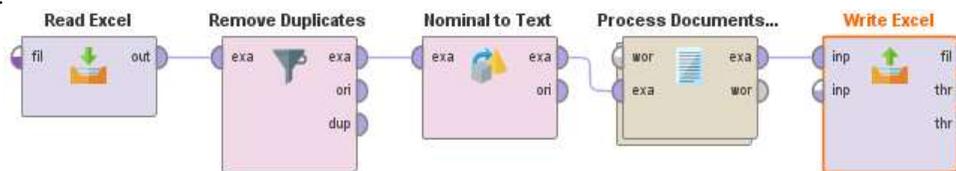
$$\text{Skor} = \sum \text{Positif} - \sum \text{Negatif}$$

$$\text{Skor} = 1 - 3 = -2$$

Menurut Aribowo (2018) Jika dalam sebuah opini mempunyai skor > 0, maka opini tersebut akan diklasifikasikan dalam kelas positif, jika opini mempunyai skor = 0 akan diklasifikasikan dalam kelas netral, sedangkan jika kalimat mempunyai skor < 0 diklasifikasikan dalam kelas negatif. Berdasarkan hasil perhitungan Tabel 2 dihasilkan skor senilai -2 yang berarti mempunyai skor < 0 dan diklasifikasikan dalam kelas negatif.

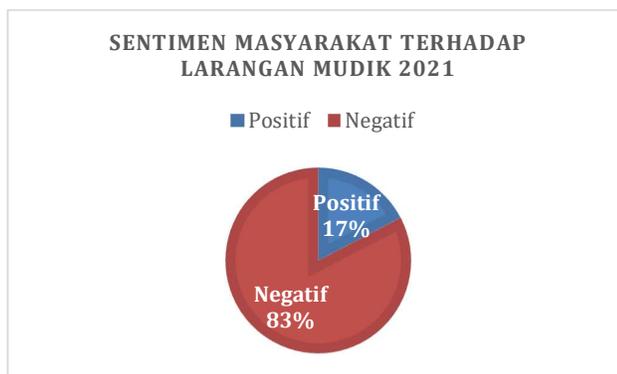
c. *Text Preprocessing*

Pada tahapan ini peneliti akan melakukan *preprocessing* pada *dataset* hal ini ditunjukkan untuk menyiapkan data yang bersih dan data yang bebas dari *noise*. Pada proses ini juga untuk menghitung pembobotan kata untuk keperluan pada proses *modelling* nanti. Berikut proses *text preprocessing* dapat dilihat pada Gambar 2.



Gambar 7. Proses *Text Preprocessing*

Setelah melewati tahapan *text preparation*, maka jumlah data *tweet* didapatkan menjadi 4.799 data yang merupakan data bersih untuk dipakai pada tahap selanjutnya, hal ini dikarenakan banyaknya data yang duplikat dan banyaknya *tweet* yang memiliki sentimen netral. Sehingga dari 4.799 data *tweet* terdapat 834 *tweet* sentimen positif dan 3.965 *tweet* sentimen negatif. Berdasarkan hal tersebut sentimen masyarakat terhadap larangan mudik 2021 lebih banyak kepada sentimen bernada negatif dengan *presentase* 83% seperti terlihat pada Gambar 3.



Gambar 8. Sentimen Masyarakat terhadap Larangan Mudik

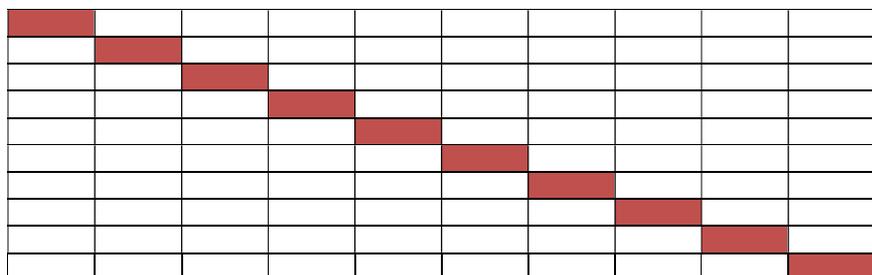
- d. Pembagian *Data Training* dan *Data Testing*
 Untuk menentukan data training dan data testing peneliti akan membagi data bersih menggunakan *split data* seperti tabel dibawah ini:

Tabel 9. Pembagian *data Training* dan *data Testing Split Data*

Algoritma	Perbandingan Data Training: Data Testing	Jumlah Data Training: Data Testing	Perbandingan jumlah positif : negatif untuk data Training	Perbandingan Jumlah positif : negatif data Testing
K-NN	60:40	2.879 : 1.920	500 : 2.379	334 : 1.586
K-NN	70:30	3.359 : 1.440	584 : 2.775	250 : 1.190
K-NN	80:20	3.839 : 960	668 : 3.172	167 : 793

Peneliti juga membagi data training dan data testing untuk k-folds validation yaitu dengan perbandingan 90:10 dengan nilai K=10. Dapat dilihat pada Tabel 4.

Tabel 10. Ilustrasi Pembagian *data Training* dan *data Testing K-fold validation*



Keterangan :

- = Kolom sebagai data *testing*
- = Kolom sebagai data *training*

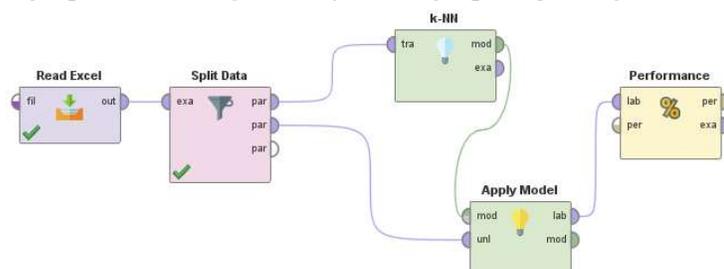
Pada Tabel 4 menunjukkan bahwa setiap percobaan terdapat 1 buah *subset* yang bergantian menjadi data *testing* dan sisanya merupakan data *training*. Pada data penelitian ini data dibagi menjadi 10 *subset* dengan jumlah 479,9 pada setiap *subset*.

3.4 Modelling

Untuk tahapan *modelling* peneliti akan melakukan pengukuran performa klasifikasi dengan *Pemodeling* menggunakan *split data*. Pemodelan ini data akan dibagi menggunakan *split* menjadi 3 model yaitu 60:40, 70:30, dan 80:20. Dan pemodelan pada *k-folds validation*.

a. *Modelling Menggunakan Split Data*

Pada tahapan ini akan dilakukan pengukuran performa klasifikasi dengan menggunakan *split data*. Berikut proses *Split data* yang dilakukan di *software RapidMiner* yang ditunjukkan pada Gambar 4.



Gambar 9. Proses *Split data* pada *RapidMiner*

Dari hasil tahapan *modelling* dengan menggunakan *Split data* maka nilai *accuracy* yang dihasilkan dapat dilihat pada Tabel 5.

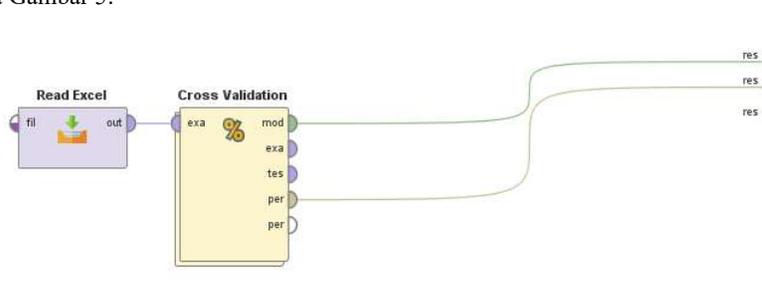
Tabel 11. Perbandingan *Accuracy* antar *Algoritma split data*

<i>Split Data</i>	K-NN
60 : 40	84.69 %
70 : 30	85.20 %
80 : 20	86.67 %

Pada Tabel 5 dihasilkan bahwa K-NN memiliki nilai akurasi tertinggi yaitu pada perbandingan *split data* 80:20 dengan nilai akurasi mencapai 86.67 %. Nilai *accuracy* merupakan gambaran seberapa akurat algoritma dapat mengklasifikasikan data secara benar. Sehingga dapat disimpulkan bahwa pembagian *split data* pada mengklasifikasikan algoritma K-NN dapat diimplementasikan dengan baik.

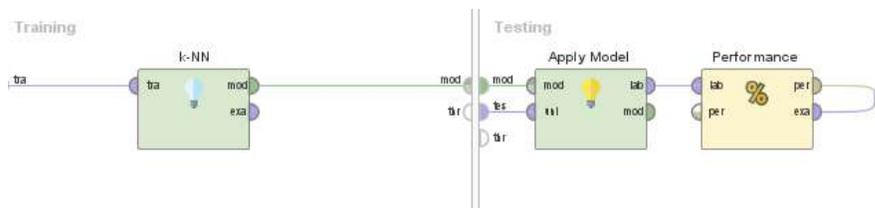
b. *Modelling Menggunakan K-Fold Cross Validation*

Pada tahapan modelling ini juga akan menggunakan *k-fold cross validation*. Hal ini peneliti lakukan untuk mengetahui apakah menggunakan *cross validation* dapat menghasilkan nilai *accuracy* yang lebih besar. contoh tahapan-tahapan modelling yang dilakukan dengan menggunakan *k-fold cross validation* dapat dilihat pada Gambar 5.



Gambar 10. Proses *Cross Validation*

6. Didalam operator *cross validation* terdapat pembagian data *training* dan data *testing* seperti pada Gambar



Gambar 11. Subprocess di *cross validation*

Setelah melakukan tahapan *modelling* menggunakan *dataset* yang telah di *preprocessing*, maka *accuracy* yang dihasilkan menggunakan *cross validation* dapat dilihat pada Tabel 6.

Tabel 12 Hasil *Accuracy K-Fold Cross Validation*

K- Fold	K-NN
2	85.16 %
3	84.37 %
4	85.08 %
5	86.24 %
6	86.52 %
7	85.74 %
8	85.97 %
9	86.56 %
10	86.56 %

Berdasarkan hasil percobaan dengan *k-fold validation* didapatkan K-NN memiliki nilai akurasi tertinggi yaitu 86.56 % pada nilai *k-fold number* = 9 dan 10. Pada setiap percobaan persentase data positif yaitu 17 % dan data negatif 83%.

Berdasarkan Tabel 5 dan Tabel 6 maka dapat disimpulkan bahwa *accuracy* tertinggi yaitu 86.67 %, yang dihasilkan dengan pemodelan menggunakan algoritma *K-Nearest Neighbor* yang menggunakan perbandingan 80:20 untuk data *training* dan data *testing*.

3.5 Evaluation

Pada penelitian ini akan dilakukan pengujian menggunakan *confusion matrix* untuk melihat hasil pengujian data yang diperoleh dari tahapan *modelling* dengan menggunakan algoritma K-NN. Perbandingan pembagian dataset yang digunakan adalah 80:20. Total *dataset* yang dikumpulkan adalah 4.799, terdapat 3.839 data *training*, dan 960 data *testing*. Berikut ini adalah *confusion matrix* hasil dari tahapan *modelling* yang dapat dilihat hasil dari perhitungan hasil *RapidMiner* dapat dilihat pada Gambar 7 untuk menunjukkan dan membuktikan hasil prediksi dari model K-NN.

accuracy: 86.67%

	true Negatif	true Positif	class precision
pred. Negatif	766	101	88.35%
pred. Positif	27	66	70.97%
class recall	96.60%	39.52%	

Gambar 12. *Confusion Matrix* pada *RapidMiner*

Perhitungan dari Gambar 7 adalah sebagai berikut :

$$a. \text{ Accuracy} = \frac{TP+TN}{TP+TN+FP+F} \times 100\% = \frac{66+766}{66+766+27+101} \times 100\% = 86.67 \%$$

$$b. \text{ Precision} = \frac{TP}{TP+FP} \times 100\% = \frac{66}{66+27} \times 100\% = 70.97 \%$$

$$c. \text{ Recall} = \frac{TP}{TP+} \times 100\% = \frac{66}{66+1} \times 100\% = 39.52 \%$$

$$d. \text{ Specificity} = \frac{TN}{TN+F} \times 100\% = \frac{766}{766+27} \times 100\% = 96.60 \%$$

Berdasarkan pada perhitungan Gambar 7 dapat dibuat kesimpulan bahwa *precision* menunjukkan tingkat ketepatan data yang diprediksi positif terhadap banyaknya data yang benar diprediksi positif yang menghasilkan *presentase* ketepatannya adalah 70.97 %, sedangkan untuk data bersentimen negatif memiliki *precision* sebesar 88.35 %. Untuk sentimen negatif memiliki *recall* (*Specificity*) adalah 96.60 % sehingga dapat disimpulkan bahwa model dapat menemukan kembali informasi atau data yang benar-benar negatif dengan sangat baik, sedangkan pada sentimen positif memiliki *recall* sebesar 36.53% karena model tidak dapat menemukan kembali data yang benar-benar positif hal ini bisa terjadi karena kurangnya data latih positif sehingga model tidak menemukan informasi kembali dengan baik. Nilai *accuracy* yang dihasilkan menggunakan model *K-Nearest Neighbor* adalah 86.67 %, sehingga dapat disimpulkan bahwa algoritma *K-Nearest Neighbor* dapat mengklasifikasi sentimen dengan baik menggunakan data Larangan Mudik 2021.

4. Kesimpulan

Berdasarkan penelitian yang telah dilakukan mengenai klasifikasi opini masyarakat Indonesia terhadap larangan mudik 2021 dapat diambil kesimpulan yaitu hasil Sentimen masyarakat terhadap larangan mudik 2021 ditemukan 17% bersentimen positif dan 83% bersentimen negatif. Maka dari hasil tersebut dapat ditarik kesimpulan bahwa masyarakat Indonesia sebagian besar memiliki respon negatif terhadap larangan mudik 2021 dan berdasarkan klasifikasi model algoritma *K-Nearest Neighbor*, menggunakan *split data* perbandingan 80:20 dengan nilai $k = 3$ terhadap *dataset* larangan mudik 2021, didapatkan nilai *accuracy* sebesar 86.67 %. Sehingga dapat dikatakan bahwa algoritma K-KNN dapat mengklasifikasikan data secara benar dan baik.

Referensi

- [1] F. Sandi dan CNBC Indonesia, "Tok! Pemerintah Resmi Larang Mudik Lebaran, Ini Alasannya," 2021. <https://www.cnbcindonesia.com/news/20210328063301-4-233390/tok-pemerintah-resmi-larang-mudik-lebaran-ini-alasannya> (diakses Apr 07, 2021).
- [2] Kominfo, "Larangan Mudik Kurangi Risiko Akibat Covid-19," www.kominfo.go.id, 2021. <https://www.kominfo.go.id/content/detail/33520/larangan-mudik-kurangi-risiko-akibat-covid-19/0/berita> (diakses Apr 05, 2021).
- [3] Tim Komunikasi Komite Penanganan Corona Virus Disease 2019 (COVID-19), "Pasien Sembuh Terus Bertambah Menjadi 1.391.742 Orang," 2021. <https://covid19.go.id/berita/pasien-sembuh-terus-bertambah-menjadi-1391742-orang> (diakses Apr 07, 2021).
- [4] Liputan6, "Terbongkar, Alasan Kuat Pemerintah Larang Mudik Lebaran 2021," 2021. <https://www.liputan6.com/bisnis/read/4524936/terbongkar-alasan-kuat-pemerintah-larang-mudik-lebaran-2021> (diakses Apr 07, 2021).
- [5] F. Sodik dan I. Kharisudin, "Analisis Sentimen dengan SVM, NAIVE BAYES dan KNN untuk Studi Tanggapan Masyarakat Indonesia Terhadap Pandemi Covid-19 pada Media Sosial Twitter," vol. 4, hal. 628–634, 2021.
- [6] T. A. M, Y. Alkhalifi, N. A. Mayangky, dan W. Gata, "Analisis Sentimen Opini Publik Mengenai Larangan Mudik pada Twitter Menggunakan Naive Bayes," *CoreIT*, vol. 6, no. 2, 2020.
- [7] R. Pebrianto, T. Rivanie, R. Nurfalalah, W. Gata, dan M. F. Julianto, "Adopsi Algorithm Support Vector Machine untuk Analisis Sentimen Larangan Mudik Lebaran 2020 pada Twitter," *J. Tek. Komput. AMIK BSI*, vol. V, no. 1, hal. 135–138, 2020, doi: 10.31294/jtk.v4i2.
- [8] R. Sari, "Analisis Sentimen Pada Review Objek Wisata Dunia Fantasi Menggunakan Algoritma K-Nearest Neighbor (K-Nn)," *EVOLUSI J. Sains dan Manaj.*, vol. 8, no. 1, hal. 10–17, 2020, doi: 10.31294/evolusi.v8i1.7371.
- [9] A. Bode, "K-Nearest Neighbor Dengan Feature Selection Menggunakan Backward Elimination Untuk Prediksi Harga Komoditi Kopi Arabika," *Ilk. J. Ilm.*, vol. 9, no. 2, hal. 188–195, 2017, doi: 10.33096/ilkom.v9i2.139.188-195.

- [10] D. T. Larose, *Discovering Knowledge in Data: An Introduction to Data Mining*, Second Edi. New Jersey: John Wiley & Son, 2014.
- [11] Herdiawan, "Analisis Sentimen Terhadap Telkom Indihome Berdasarkan Opini Publik Menggunakan Metode Improved K-Nearest Neighbor," *J. Ilm. Komput. dan Inform.*, 2018.
- [12] F. Z. Tala, "A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia," *M.Sc. Thesis, Append. D*, vol. pp, hal. 39-46, 2003.
- [13] D. H. Wahid dan A. SN, "Peringkasan Sentimen Esktraktif di Twitter Menggunakan Hybrid TF-IDF dan Cosine Similarity," *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 10, no. 2, hal. 207, 2016, doi: 10.22146/ijccs.16625.