

ANALISIS SENTIMEN OPINI PUBLIK TERHADAP EFEK PSBB PADA TWITTER DENGAN ALGORITMA DECISION TREE-KNN-NAÏVE BAYES

Muhammad Syarifuddin

Pendidikan Teknologi Informasi

Universitas Brawijaya

www.ub.ac.id

syariff@student.ub.ac.id/syarifuddinm79@gmail.com

Abstract- *Community aspirations are sometimes difficult to convey to the person in charge directly, it encourages people to express their aspirations, criticisms and the like through social media, one of which is the popular social media today is Twitter. One collection of aspirations or tweets from Twitter users about the PSBB effect, one of which, can be used as an analysis of public opinion sentiments. Data on the effects of PSBB were obtained as many as 170 opinions, then processed using data mining techniques (data mining), in which there are processes of text mining, tokenize, transformation, classification, and stem. Then it is calculated into three different algorithms to be compared, the algorithm used is Decision Tree, K-Nearest Neighbors (K-NN), and Naïve Bayes Classifier with the aim of finding the best accuracy. Rapidminer Application Version 7.1 is also used to facilitate writers in processing data. The highest results from this study were the Decision Tree algorithm with an accuracy value of 83.3%, precision 79% and recall 87.17%.*

Key words: *twitter, data mining, rapidminer, sentiment analysis, algorithms.*

Abstrak- Aspirasi masyarakat terkadang sulit untuk disampaikan kepada pihak penanggung jawab secara langsung, hal tersebut mendorong masyarakat untuk menyampaikan aspirasi, kritik dan sejenisnya melalui media sosial, salah satunya media sosial yang populer saat ini adalah *twitter*. Kumpulan aspirasi atau *tweet* dari pengguna *twitter* mengenai efek PSBB salah satunya, dapat dimanfaatkan menjadi sebuah analisis sentimen opini publik. Data mengenai efek PSBB didapatkan sebanyak 170 opini, kemudian diolah menggunakan teknik penambangan data (*data mining*), didalamnya terdapat proses penambangan teks, *tokenize*, transformasi, klasifikasi, dan *stem*. Kemudian dikalkulasikan kedalam tiga algoritma yang berbeda untuk dibandingkan, algoritma yang digunakan yaitu *Decision Tree*, *K-Nearest Neighbors* (K-NN), dan *Naïve Bayes Classifier* dengan tujuan menemukan akurasi terbaik. Aplikasi *Rapidminer Versi 7.1* juga digunakan untuk mempermudah penulis dalam mengolah data. Hasil tertinggi dari penelitian ini adalah algoritma *Decision Tree* dengan nilai *accuracy* 83,3%, *precision* 79% dan *recall* 87,17%.

Kata Kunci: *twitter, data mining, rapidminer, analisis sentimen, algoritma.*

PENDAHULUAN

Saat ini sebagian besar pengguna jalan dan kendaraan sedang melaksanakan PSBB (Pembatasan Sosial Berskala Besar). PSBB merupakan kebijakan yang pertama kali diusulkan oleh Anies Baswedan yang saat ini menjabat sebagai Gubernur DKI Jakarta (Irani, 2020). Sementara itu salah satu kendaraan yang paling sering digunakan adalah kendaraan pribadi roda dua maupun roda 4. Di sisi lain, penggunaan sosial media semakin meningkat, adapun penggunaan jejaring internet saat ini dimanfaatkan masyarakat luas untuk menyampaikan aspirasi dan pendapatnya secara bebas, salah satunya menggunakan sosial media yang dimiliki (Septiani & Sibaroni, 2019). *Twitter*, merupakan salah satu sosial media yang sering digunakan masyarakat

dalam beraspirasi (Rafita, 2014). *Twitter* dianggap sebagai sosial media yang mudah untuk digunakan dan sangat cepat dalam menyebarkan informasi. saat ini masyarakat dalam menggunakan *twitter* tidak hanya untuk kepentingan pribadi saja, melainkan digunakan juga untuk kepentingan bisnis maupun politik (Septiani & Sibaroni, 2019). Perusahaan atau instansi terkait membutuhkan pendapat atau aspirasi dari pengguna *twitter* secara umum untuk tujuan tertentu, salah satunya dengan melakukan analisis sentimen terhadap pendapat umum tersebut (Septiani & Sibaroni, 2019).

Pendapat para pengguna jalan dan kendaraan melalui *twitter* dapat menyimpulkan kondisi yang sedang terjadi di wilayah yang sedang diadakan PSBB. Dikarenakan banyaknya komentar yang ada di *twitter* dan bahkan sering kali ada topik yang

menjadi *trending* di *twitter*. karena hal itulah penulis merasa perlu untuk menganalisis sentimen pada *twitter* tentang efek PSBB.

Di sisi lain, *data mining* adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode atau algoritma *data mining* sangat bervariasi (Muzakir & Wulandari, 2016). Dengan demikian *data mining* dapat digunakan sebagai analisis sentimen karena dapat mengolah data *tweet* dalam jumlah banyak.

Di sisi lain, menurut (Muljono et al., 2018), analisis sentimen merupakan sebuah proses penemuan teks yang diungkapkan pengguna. Dalam konteks penelitian ini yaitu menentukan teks dalam *twitter* apakah teks tersebut bermakna negatif, positif atau bahkan bernilai netral.

Penelitian ini membahas bagaimana melakukan analisis sentimen menggunakan opini masyarakat Indonesia di *twitter*, nantinya opini tersebut diklasifikasikan menjadi data opini sentimen positif dan negatif. Penelitian ini menggunakan tiga algoritma berbeda yaitu, *Decision Tree*, *K-NN*, dan *Naïve Bayes* untuk mengekstrak pemikiran atau perasaan pengguna melalui *tweet* mereka dan mengelompokkan datanya ke dalam kategori yang berbeda. Adapun tujuan dari penelitian ini yaitu, untuk membandingkan hasil dari ketiga algoritma tersebut untuk mengetahui *classifier* mana yang memberikan hasil terbaik dalam hal *accuracy*, *precision*, dan *recall*. Sehingga penelitian ini dapat menjadi dasar penelitian yang akan datang dengan metode yang paling tepat.

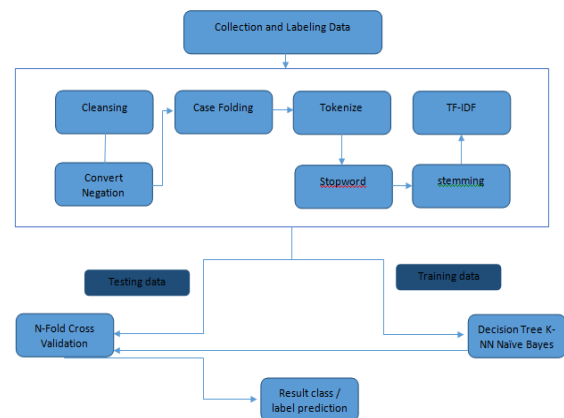
BAHAN DAN METODE

Penelitian tentang analisis sentimen dari *tweet* di *twitter* tentang pemberlakuan PSBB di Indonesia dalam pencegahan penyebaran COVID-19. Secara umum, suatu analisis sentimen digunakan untuk memprediksi kecenderungan suatu opini masyarakat terhadap sesuatu hal, apakah cenderung ke opini positif atau negatif (Rozi et al., 2012). Apa yang penulis paparkan tadi adalah bukti bahwa sebelumnya sudah ada penelitian mengenai analisis sentimen dari penggunaan media sosial *twitter*.

Analisis sentimen dari *twitter* harus fokus pada masalah klasifikasi. Seperti yang diungkapkan Rani (2015), Klasifikasi merupakan suatu proses pencarian suatu fungsi atau himpunan sehingga menjadi suatu konsep yang utuh, dengan tujuan dapat menggunakan model tersebut untuk memprediksi kelas dari suatu objek yang mana kelasnya belum diketahui (Rani, 2015). Berbagai pendekatan klasifikasi seperti *Naïve Bayes*, *Decision Tree*, dan *K-Nearest Neighbor* (K-NN) telah diterapkan untuk menemukan hasil terbaik.

Dengan demikian, penambangan dapat digunakan untuk tingkat analisis sentimen. *Data mining* adalah proses menemukan pola atau pengetahuan. Pola harus *valid*, berguna, dan dapat dipahami dalam sejumlah langkah: pra-pemrosesan, penambangan data, dan pasca-pemrosesan.

Dalam menganalisis sentimen, ada beberapa tahapan yang perlu dilakukan untuk mendapatkan hasil tes terbaik. Adapun Gambar 1 menunjukkan langkah-langkah yang penulis gunakan yang dimulai dari *collection and labeling data* hingga *label prediction*.

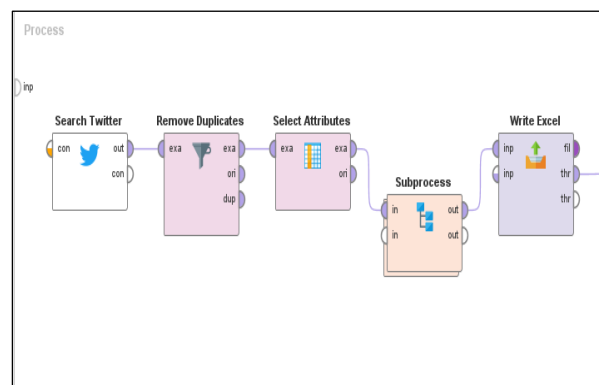


Sumber: (Syarifuddin, 2020b)

Gambar 1. Diagram Sistem Klasifikasi Analisis Sentimen

Pengumpulan dan Pelabelan Data

Tahap pertama dalam melakukan proses analisis sentimen adalah pengumpulan data. Data diambil dari *twitter* dengan permintaan pencarian tentang “Efek PSBB” sebanyak 170 catatan menggunakan aplikasi *Rapidminer*.



Sumber: (Syarifuddin, 2020b)

Gambar 2. Pengambilan data dari *Twitter*

Search twitter digunakan untuk pengambilan data dari sosial media *twitter* yang dilanjut dan disimpan dalam bentuk *excel* dengan *write excel*. Dilanjut dengan konten *remove duplicates* yang

berfungsi untuk menghapus konten yang berganda atau duplikat. Dilanjut dengan proses pemberian label pada konten, yaitu untuk mengelompokkan data menjadi beberapa kelas yang akan digunakan oleh penulis, adapun kelas yang akan digunakan adalah data bernilai positif dan negatif. Sedangkan tujuan dari pelabelan ini sendiri yaitu untuk membagi data menjadi 2 set, yaitu data *testing* dan data *training*. Adapun data *testing* merupakan data yang akan penulis uji cobakan terhadap data *training*, sedangkan data *training* adalah data yang digunakan untuk menggali informasi dari sistem agar suatu pola dapat dikenali (Rasenda et al., 2020). Adapun lebih jelas terdapat pada Gambar 2 diatas dan berikut ini adalah salah satu contoh *dataset* yang telah diberi label:

Tabel 1. Contoh Label Dataset

Text	Sentiment
Efek psbb kemaren berasa banget ke udara jakarta yg bikin langitnya bersih, skrang udh new normal plis yg mulai beraktifitas kita jaga lagi udaranya, terutama yg pake kendaraan soalnya ngaruh bgt asepe kendaraan sama kualitas udara kita.	POSITIVE
Halusinasi yg berkepanjangan. Efek PSBB kelamaan.	NEGATIVE

Sumber: (Syarifuddin, 2020a)

Dan berikut contoh data yang diambil dari www.twitter.com:



Sumber: (@king_aster, 2020)

Gambar 3. Opini Negatif Efek PSBB



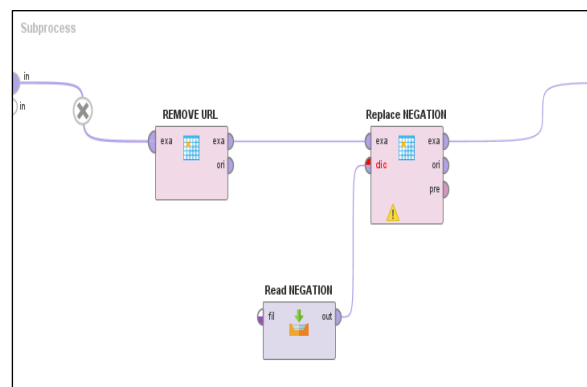
Sumber : (@Naunaaty, 2020)

Gambar 4. Opini Positif Efek PSBB

HASIL DAN PEMBAHASAN

A. Pra-Pemrosesan

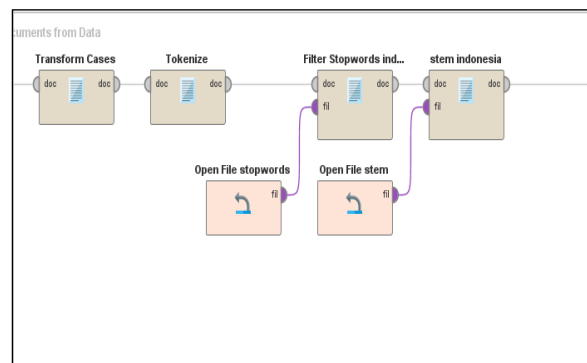
Setelah data diberi label, langkah selanjutnya adalah pra-pemrosesan. Tahap ini adalah tahap di mana data dipersiapkan untuk menjadi data yang siap dianalisis. Ada beberapa tahapan dalam *preprocessing* ini, termasuk *cleansing*, *convert negation*, *case folding*, *tokenization*, *filtering stopword* dan *stemming* dalam bahasa Indonesia. Gambar 5 menunjukkan operator "Subprocess". Dalam hal ini menggunakan konten "Remove URL", "Convert Negation".



Sumber: (Syarifuddin, 2020b)

Gambar 5. Operator Subprocess

Gambar 6 menunjukkan operator "Process Document", menggunakan konten "Transform Cases", "Tokenize", "Stopword Filter" dan "Stem" dalam bahasa Indonesia.



Sumber: (Syarifuddin, 2020b)

Gambar 6. Isi Operator Process Document

Berikut ini adalah deskripsi terperinci dari tahapan pra-pemrosesan di atas:

1. Cleansing

Tahap proses ini adalah proses untuk membersihkan atribut yang tidak diperlukan dalam input data seperti simbol dan tanda baca (Amalia et al., 2019). Bekerja untuk mengurangi *noise* pada *dataset*. Contoh karakter yang

dihilangkan seperti tanda baca seperti titik (.), Koma (,) dan tanda baca lainnya. Berikut ini adalah contoh kalimat pembersihan data, *input* “Efek psbb rasa *lockdown* langsung rugi triliunan <https://t.co/K6uMYZG5vc>”, *output* “Efek psbb rasa *lockdown* langsung rugi triliunan”.

2. Convert Negation

Atau yang dikenal dengan kata negasi, merupakan suatu makna kata yang sesungguhnya yang dapat dibalikkan. Berikut ini adalah contoh negasi konversi kalimat, *input* “Efek PSBB rasa *lockdown* langsung rugi triliunan”, *output* “Efek psbb rasa *lockdown* langsung rugi triliunan”.

3. Case Folding

Dalam menulis *tweet*, biasanya ada bentuk huruf yang berbeda, tahap ini adalah proses keseragaman huruf, berkaitan dengan huruf kapital atau bukan. Berikut ini adalah contoh kalimat *case folding*, *input* “Efek PSBB rasa *lockdown* langsung rugi triliunan”, menjadi *output* “efek psbb rasa *lockdown* langsung rugi triliunan”.

4. Tokenization

Menurut Rasenda (2020), menjelaskan bahwa *tokenization* merupakan proses memilah dan memisahkan suatu kalimat menjadi beberapa kata, yang disebut dengan token. Dalam kalimat tertentu, proses ini juga dapat menghilangkan suatu tanda baca yang tidak diperlukan sehingga akan mempermudah dalam proses olah data pada *rapidminer* (Rasenda et al., 2020). Terdapat model *tokenization* yang dapat digunakan diantaranya yaitu, *unigram*, *bigram*, *trigram*, dan *ngram*. Berikut ini adalah contoh kalimat *tokenization*, *input* "efek psbb rasa *lockdown* langsung rugi triliunan", *output* "efek, psbb, rasa, *lockdown*, langsung, rugi, triliunan".

5. Filtering

Filtering adalah tahap menghilangkan kata-kata yang muncul dalam jumlah besar tetapi dianggap tidak memiliki makna (*stopwords*). Pada dasarnya, daftar *stopwords* adalah sekumpulan kata yang banyak digunakan dalam berbagai bahasa. Alasan menghapus kata yang terkait dengan penambahan teks adalah karena penggunaannya yang terlalu umum, sehingga pengguna dapat fokus pada kata-kata lain yang jauh lebih penting. Berikut ini adalah contoh kalimat *stopwords*, *input* “efek psbb rasa *lockdown* langsung rugi triliunan”, *output* “efek psbb langsung rugi triliunan”. Berikut adalah salah contoh kata dari *stopwords*:

Tabel 2. Contoh Kata *Stopword*

Rasa	Lockdown	kalau	pada	yaitu
Aku	Dia	kami	saja	asik
bapak	Ini	lalu	hai	hari
berbagai	Itu	lewat	untuk	masa
Cara	jadi	meski	yang	tapi
Cuma	juga	oleh	wah	hal

Sumber: (Pertiwi, 2019)

6. Stemming

Stemming adalah proses pemetaan dan penguraian berbagai bentuk (*variants*) dari suatu kata menjadi bentuk kata dasarnya. Algoritma *stemming* dikembangkan berdasarkan aturan morfologi Bahasa Indonesia, yang mengelompokkan imbuhan menjadi awalan (*prefix*), sisipan (*infix*), akhiran (*suffix*) dan gabungan awalan akhiran (*confixes*). Adapun algoritma ini, menyusun kembali kata-kata yang berlebihan dalam suatu proses *stemming*, atau yang disebut *recoding* dan menggunakan suatu kamus kata dasar (Wahyudi et al., 2017). Berikut ini adalah contoh dari kalimat, *input* “efek psbb rasa *lockdown* langsung rugi triliunan”, dan *output* “efek psbb rasa *lockdown* langsung rugi triliunan”. Berikut ini adalah salah satu contoh kata dari *stemming*:

Tabel 3. Contoh Kata *Stemming*

Sebelum	Sesudah
Triliunan	Triliun
Serendah	Rendah
Sebelum	Belum
Diberikan	Beri
Secukupnya	cukup
Dilaksanakan	laksanakan

Sumber: (Pertiwi, 2019)

7. Weighting Word

Weighting word adalah mekanisme memberi skor kemunculan kata dalam dokumen teks dan untuk menghitungnya dapat menggunakan Persamaan 1 dibawah ini (Rasenda, 2020)

$$tf - idf_{t,d} = tf_d * idf_t \dots\dots\dots(1)$$

Frekuensi istilah (tf) adalah frekuensi kemunculan istilah (t) dalam dokumen (d) Salah satu metode populer untuk menimbang kata adalah TF-IDF. Metode TF-IDF menurut Meliana (2008), merupakan metode yang menggabungkan dua konsep untuk perhitungan bobot, yaitu frekuensi kemunculan sebuah kata yang berada dalam suatu *inverse* frekuensi dan dokumen

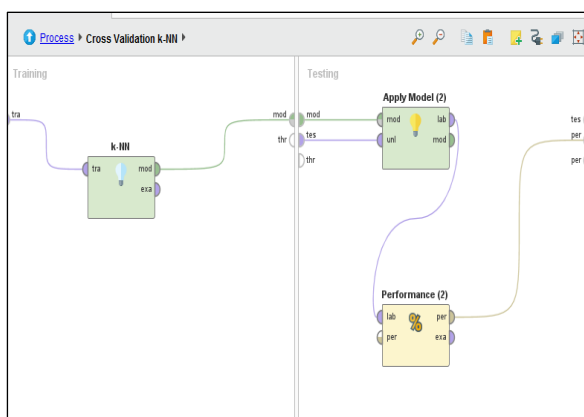
tertentu yang didalamnya terdapat kata tersebut (Meliana, 2008). Frekuensi kemunculan kata di dalam dokumen yang diberikan menunjukkan seberapa penting kata itu di dalam dokumen tersebut. Frekuensi dokumen yang mengandung kata tersebut menunjukkan seberapa umum kata tersebut. Jika frekuensi kata dalam dokumen tersebut memiliki nilai tinggi, sedangkan nilai frekuensi keseluruhan dokumen rendah, maka nilai bobot hubungan antara sebuah dokumen dan kata pasti tinggi (Sugiyanto & Nugraha Putra, 2013).

Sedangkan Frekuensi Dokumen adalah jumlah dokumen tempat *term* muncul. Semakin kecil frekuensi kemunculannya, semakin kecil nilainya. Saat menghitung frekuensi istilah, semua kata di dalamnya dianggap penting. Namun, ada kata-kata yang sebenarnya kurang penting dan tidak perlu diperhitungkan seperti "di-", "kem", "dan", dll.

Oleh karena itu, kata-kata yang kurang penting ini perlu dikurangi dan menambah dengan kata-kata penting lainnya. Ini adalah ide dasar mengapa *stopword* dibutuhkan. Dengan demikian diperlukan TF-IDF untuk menghitung, skor dapat diperoleh dengan menggunakan Persamaan. Metode TF-IDF menentukan frekuensi relatif kata-kata dalam dokumen tertentu melalui proporsi terbalik kata di atas seluruh kumpulan dokumen (Trstenjak et al., 2014).

B. Klasifikasi Analisis Sentimen

Setelah pra-pemrosesan data, langkah selanjutnya adalah klasifikasi analisis sentimen. Tahap ini adalah tahap untuk memberikan pelatihan dan mengimplementasikan berbagai algoritma penambangan data. Gambar 7 menunjukkan konten operator "Cross Validation" dalam aplikasi *Rapidminer*. Dalam hal ini, menggunakan tiga operator klasifikasi berbeda untuk perbandingannya, yaitu operator klasifikasi *Decision Tree*, *K-NN* dan *Naïve Bayes*.



Sumber: (Syarifuddin, 2020b)

Gambar 7. Cross Validation Operator

C. Evaluasi Analisis Sentimen

Setelah proses pengujian ini selesai dilakukan, diperlukan satu tahap lagi untuk menentukan kualitas satu proses yang sudah dilakukan, yaitu evaluasi hasil. Evaluasi hasil dilakukan dengan tiga parameter yaitu, *precision*, *accuracy* dan *recall*, dari suatu perhitungan yang sudah dilakukan (Hadna et al., 2016).

Akurasi (A) adalah jumlah dokumen yang diklasifikasikan dengan benar, baik Benar Positif maupun Benar Negatif. Menghitung nilai akurasi dapat menggunakan Persamaan 2:

$$A = \frac{TP+TN}{(TP+FP+TN+FN)} \times 100\% \dots\dots\dots(2)$$

Presisi (P) adalah seberapa banyak hasil pemrosesan relevan dengan informasi yang ingin dicari. Dengan kata lain, presisi adalah klasifikasi *True Positive* dan semua data diprediksi sebagai kelas positif. Menghitung nilai presisi dapat menggunakan Persamaan 3:

$$P = \frac{TP}{(TP+FP)} \times 100\% \dots\dots\dots(3)$$

Recall (R) adalah berapa banyak dokumen yang relevan dalam koleksi dihasilkan oleh sistem. Dengan kata lain, *recall* adalah jumlah dokumen yang memiliki klasifikasi Benar Positif dari semua dokumen yang benar-benar positif (termasuk *False Negative*). Menghitung nilai *recall* dapat menggunakan Persamaan 4:

$$P = \frac{TP}{TP+FN} \times 100\% \dots\dots\dots(4)$$

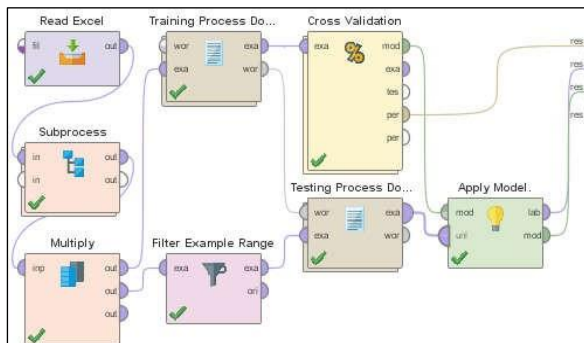
Variabel seperti TN, TP, FN dan FP berasal dari *confusion matrix*. TN adalah singkatan dari *True Negative*, data negatif yang diklasifikasikan sebagai negatif. Sedangkan *True Positive* merupakan kepanjangan dari TP, yaitu suatu data positif yang akan diklasifikasikan menjadi positif. *False Negative* merupakan kepanjangan dari FN, yaitu suatu data yang positif kemudian akan diklasifikasikan menjadi data negatif. FP adalah singkatan dari *False Positive*, data negatif yang diklasifikasikan sebagai positif (Hadna et al., 2016). Untuk penjelasan yang lebih detail terdapat pada Tabel 4.

Tabel 4. Confusion Matrix

	Prediction Yes	Prediction No
True Yes	TP	FN
True No	FP	TN

Sumber: (Rasenda, 2020)

Setelah data dikumpulkan, data akan dibagi menjadi data *training* dan data *testing*. Adapun metode *N-fold cross* digunakan dalam pembagian data dan menghilangkan suatu bias data. Metode ini membagi suatu dokumen menjadi *n* bagian. Dalam berbagai eksperimen, adapun yang akan dilakukan oleh penulis yaitu, sebagai pengujian data, $(n-1)/2$ bagian lain sebagai dokumen tanpa label dan $(n-1)/2$ bagian sebagai dokumen berlabel, yang akan ditukar sebanyak *n* kali dalam setiap eksperimen (Rasenda et al., 2020). Kumpulan dokumen yang dimiliki pertama secara acak diurutkan sebelum dimasukkan ke dalam lipatan. Hal ini dilakukan untuk menghindari pengelompokan dokumen dari satu kategori tertentu di lipatan. Dilanjut dengan penjelasan hasil-hasil eksperimen dan analisis kinerja



Sumber: (Syarifuddin, 2020b)

Gambar 8. Proses Utama Pada Rapidminer

Pada Gambar 8 menunjukkan proses utama dalam aplikasi Rapidminer. Konten "Read Excel" digunakan untuk membaca data dalam file Excel. Konten "Subprocess" dan "Process Document" digunakan untuk pra-pemrosesan. Adapun eksperimen dilakukan sebanyak 10 kali dengan bantuan konten *cross validation* guna evaluasi hasil analisis serta hasil klasifikasi (Rasenda et al., 2020).

Name	Type	Missing	Filter (1,862 / 1,862 attributes)	Search for Attributes
Label PREDICATED CLASS	Binominal	0	Label POSITIVE (111)	Label NEGATIVE (111)
a	Real	0	Min 0	Max 0.301
aaaaa	Real	0	Min 0	Max 0.162
aamiin	Real	0	Min 0	Max 0.324
abis	Real	0	Min 0	Max 0.199
acan	Real	0	Min 0	Max 0.151
acuan	Real	0	Min 0	Max 0.173

Sumber: (Syarifuddin, 2020b)

Gambar 9. Hasil Prediksi Rapidminer

Gambar 9 menunjukkan hasil prediksi dalam aplikasi Rapidminer. Melihat sentimen dan juga label prediksi yang sama dengan label sentimen.

Berikut ini gambaran hasil dari *confusion matrix* dari masing-masing algoritma pada Rapidminer:

Tabel 5. Confusion Matrix Setiap Algoritma

Metode	TP	FP	TN	FN
Decision Tree	34	9	36	5
K-NN	32	8	48	11
Naive Bayes	37	6	47	22

Sumber: (Syarifuddin, 2020b)

Dari *confusion matrix* pada Tabel 5 nilai rata-rata *accuracy*, *precision* dan *recall* ditampilkan dengan perhitungan menggunakan rumus (8), (9), (10).

Tabel 6. Nilai Accuracy, Precision, dan Recall Menggunakan Formula

Metode	Accuracy	Precision	Recall
Decision Tree	83,3 %	81,06 %	87,17 %
K-NN	80,80 %	82,72 %	74,41 %
Naive Bayes	80,03 %	87,54 %	62,71 %

Sumber:

Ada perbedaan dalam hasil nilai rata-rata presisi menggunakan aplikasi Rapidminer seperti yang ditunjukkan pada Tabel 7 di bawah ini:

Tabel 7. Nilai Accuracy, Precision, dan Recall Menggunakan Rapidminer

Metode	Accuracy	Precision	Recall
Decision Tree	83,3 %	79%	87,17 %
K-NN	80,80 %	80%	74,41 %
Naive Bayes	80,03 %	86%	62,71 %

Sumber: (Syarifuddin, 2020b)

Hasil ini menunjukkan *accuracy* dari Decision Tree, KNN, dan Naive Bayes sebesar 83,3%, 80,80%, dan 80,03%. Hasil untuk *precision* dari Decision Tree, K-NN, dan Naive Bayes sebesar 81,06%, 82,72%, dan 87,54%. Sementara hasil untuk *Recall* dari Decision Tree, K-NN, dan Naive Bayes adalah 87,17%, 74,41%, dan 62,71%. Sehingga dapat disimpulkan bahwa *classifier Decision Tree* adalah *classifier* terbaik untuk digunakan dengan dataset media sosial karena memberikan prediksi yang lebih akurat dan tepat.

KESIMPULAN

Dalam studi ini, upaya dilakukan untuk mengklasifikasikan sentimen analisis dari tweet di twitter tentang efek PSBB bagi masyarakat di Indonesia. Untuk merangkul pandangan

masyarakat ini teknik penambangan teks digunakan, dan penggalian data menggunakan tiga algoritma yang berbeda yaitu *Decision Tree*, K-NN, dan *NaiveBayes*. Tiga algoritma memprediksi label pada *dataset*. Dari hasil pengujian secara keseluruhan algoritma *Decision Tree* memiliki nilai tertinggi dari nilai *accuracy* dan *recall*. Sehingga dapat ditarik benang merah bahwa algoritma *Decision Tree* adalah *classifier* terbaik dalam memberikan prediksi yang lebih akurat dan tepat tentang efek PSBB di Indonesia.

REFERENSI

- @king_aster. (2020). *Halusinasi yg berkepanjangan. Efek PSBB kelamaan.* [https://twitter.com/search?q=Halusinasi yg berkepanjangan&src=typeahead_click](https://twitter.com/search?q=Halusinasi%20yg%20berkepanjangan&src=typeahead_click)
- @Naunaaty. (2020). *Efek psbb kemaren berasa banget ke udara jakarta yg bikin langitnya bersih, skrang udh new normal plis yg mulai beraktifitas kita jaga lagi udaranya, terutama yg pake kendaraan soalnya ngaruh bgt asepe.* Twitter.com. [https://twitter.com/search?q=Efek psbb kemaren berasa banget ke udara jakarta yg bikin langitnya bersih%2C skrang udh new normal plis yg mulai beraktifitas kita jaga lagi udaranya%2C terutama yg pake kendaraan soalnya ngaruh bgt asepe &src=typeahead_click](https://twitter.com/search?q=Efek%20psbb%20kemaren%20berasa%20banget%20ke%20udara%20jakarta%20yg%20bikin%20langitnya%20bersih%20skrang%20udh%20new%20normal%20plis%20yg%20mulai%20beraktifitas%20kita%20jaga%20lagi%20udaranya%20terutama%20yg%20pake%20kendaraan%20soalnya%20ngaruh%20bgt%20asepe&src=typeahead_click)
- Amalia, A., Gunawan, D., Fithri, Y., & Aulia, I. (2019). Automated Bahasa Indonesia essay evaluation with latent semantic analysis. *Journal of Physics: Conference Series*, 1235(1). <https://doi.org/10.1088/1742-6596/1235/1/012100>
- Hadna, M. S., Santosa, P. I., & Winarno, W. W. (2016). Studi Literatur Tentang Perbandingan Metode Untuk Proses Analisis Sentimen Di Twitter. *Seminar Nasional Teknologi Informasi Dan Komunikasi, 2016*(Sentika), 57–64.
- Irani, W. M. A. (2020). Breaking News - PSBB DKI Jakarta, Anies_ Akan Disusun Peraturan yang Mengikat. *Tribunnews.Com*, Corona.
- Meliana, N. (2008). Deteksi Kesesuaian Bidang Minat Terhadap Proposal Tugas Akhir Mahasiswa. *Ukdw*.
- Muljono, Artanti, D. P., Syukur, A., Prihandono, A., & Setiadi, D. R. I. M. (2018). Analisa Sentimen Untuk Penilaian Pelayanan Situs Belanja Online Menggunakan Algoritma Naive Bayes. *Konferensi Nasional Sistem Informasi*, 8–9.
- Muzakir, A., & Wulandari, R. A. (2016). Model Data Mining sebagai Prediksi Penyakit Hipertensi Kehamilan dengan Teknik Decision Tree. *Scientific Journal of Informatics*, 3(1), 19–26. <https://doi.org/10.15294/sji.v3i1.4610>
- Pertiwi, M. W. (2019). Analisis Sentimen Opini Publik Mengenai Sarana Dan Transportasi Mudik Tahun 2019 Pada Twitter Menggunakan Algoritma Naive Bayes, Neural Network, KNN dan SVM. *Inti Nusa Mandiri*, 14(1), 27–32.
- Rafita, Y. (2014). SOCIAL NETWORK ANALYSIS DALAM MELIHAT KECENDERUNGAN PEMBERITAAN PADA AKUN TWITTER “@detikcom” dan “@Metro_TV.” *Khazanah*, 6(2), 67–81. <https://doi.org/10.20885/khazanah.vol6.iss2.art7>
- Rani, L. N. (2015). Klasifikasi Nasabah Menggunakan Algoritma. *Jurnal KomTekInfo Fakultas Ilmu Komputer*, 2(2), 33–38.
- Rasenda. (2020). Implementasi K-NN Dalam Analisa Sentimen Riba Pada Bunga Bank Berdasarkan Data Twitter. *Jurnal Informatika*, 7(April), 1–8. <https://doi.org/10.30865/mib.v4i2.2051>
- Rasenda, R., Lubis, H., & Ridwan, R. (2020). Implementasi K-NN Dalam Analisa Sentimen Riba Pada Bunga Bank Berdasarkan Data Twitter. *Jurnal Media Informatika Budidarma*, 4(2), 369. <https://doi.org/10.30865/mib.v4i2.2051>
- Rozi, I. F., Pramono, S. H., & Dahlan, E. A. (2012). Implementasi Opinion Mining (Analisis Sentimen) untuk Ekstraksi Data Opini Publik pada Perguruan Tinggi. *Electrical Power, Electronics, Communications, Controls, and Informatics Seminar (EECCIS)*, 6(1), 37–43.
- Septiani, L., & Sibaroni, Y. (2019). Sentiment Analysis Terhadap Tweet Bernada Sarkasme Berbahasa Indonesia. *Jurnal Linguistik Komputasional (JLK)*, 2(2), 62. <https://doi.org/10.26418/jlk.v2i2.23>
- Sugiyanto, S., & Nugraha Putra, W. H. (2013). Identifikasi Duplikasi Laporan Bug Pada Repositori Laporan Bug Untuk Menghasilkan Saran Resolusi Bug Perangkat Lunak. *JUTI: Jurnal Ilmiah Teknologi Informasi*, 11(2), 35.

<https://doi.org/10.12962/j24068535.v11i2.a9>

Syarifuddin, M. (2020a). *Hasil Pencarian Pada Twitter Dengan Kata Kunci "COVID-19."* Twitter.Com.

https://twitter.com/search?q=COVID-19&src=typed_query

Syarifuddin, M. (2020b). *Laporan Akhir Penelitian Mandiri: Analisis Sentimen Opini Publik Mengenai Covid-19 Pada Twitter Menggunakan Metode Naïve Bayes Dan Knn.*

Trstenjak, B., Mikac, S., & Donko, D. (2014). KNN with TF-IDF based framework for text categorization. *Procedia Engineering*, 69, 1356–1364.

<https://doi.org/10.1016/j.proeng.2014.03.129>

Wahyudi, D., Susyanto, T., & Nugroho, D. (2017). Implementasi Dan Analisis Algoritma Stemming Nazief & Adriani Dan Porter Pada Dokumen Berbahasa Indonesia. *Jurnal Ilmiah SINUS*, 15(2).

<https://doi.org/10.30646/sinus.v15i2.305>