

ANALISIS SENTIMEN KULIAH DARING DENGAN ALGORITMA NAÏVE BAYES, K-NN DAN DECISION TREE

EI Miana Assni Ernania¹, Asti Herliana²

¹Universitas Adhirajasa Reswara Sanjaya
email: eassnie@gmail.com

²Universitas Adhirajasa Reswara Sanjaya
email: asti@ars.ac.id

Abstrak

Kuliah daring dianggap sebagai solusi untuk terus menjalankan proses mengajar dimasa pandemi. Namun kurangnya adaptasi dan perubahan secara mendadak menyebabkan berbagai respon dan opini masyarakat bermunculan dimedia sosial. Hal ini menimbulkan sikap pro dan kontra dari berbagai pihak. Untuk itu penelitian ini bertujuan untuk melakukan penambangan text di twitter. Penelitian ini dilakukan dengan menggunakan algoritma *Naïve Bayes*, *K-Nearest Neighbor* dan *Decision Tree* untuk membandingkan tingkat akurasi dari ketiga algoritma tersebut. Hasil penelitian menunjukkan bahwa analisis sentimen terhadap data tweet kuliah daring dengan menggunakan naïve bayes dengan teknik *cross validation* memperoleh hasil sebesar 81,57%. Untuk class precision prediksi positif 100%, prediksi negative memperoleh hasil 73,06% dan recall yang diperoleh dari true positif sebesar 63,13% sedangkan untuk *true negative* sebesar 100%. Lalu algoritma *K-Nearest Neighbor* menghasilkan nilai *accuracy* 62,10% untuk *class precision* prediksi positif adalah 62,06% sedangkan untuk *prediksi negative* memperoleh hasil 62,13% dan *recall* yang diperoleh dari true positif sebesar 62,24% sedangkan untuk *true negative* sebesar 61,95% dan hasil dari algoritma *Decision tree* memperoleh nilai *accuracy* 51,60% untuk class precision prediksi positif adalah 51,89% sedangkan untuk prediksi negative memperoleh hasil 51,31% dan *recall* yang diperoleh dari *true positif* sebesar 51,74% sedangkan untuk *true negative* sebesar 51,47%.

Kata Kunci: *Decision Tree, K-NN, Kuliah daring, Naive bayes, Twitter, Text mining*

Abstract

Online lectures have been deemed a solution to continuing the teaching process at pandemic. But the lack of adaptation and sudden change has resulted in public responses and opinions arising on social media. This has created a wide range of controversy. To that end the research aims to deliver text mining on twitter. The study uses the naive bayes algorithm, k-nearest neighbor and decision tree to compare the levels of accuracy of the three methods. Research has shown that sentient analysis of the online college's twitter data by using naive bayes with a cross validation technique gets 81.57% results. For class precision's positive predictions are 100%, negative predictions get 73.06% results and recall obtained from 63.13% true positive whereas for true negative is 100%. Then the k-nearest neighbor algorithm yields a benefit 62.06% for a class precision positive prediction, whereas for negative predictions get 62.13% and a recall obtained from a 62.24% true positive and 61.60% for a true negative for 61.1.60% for a class precision positive predictions is 51.89% while for a negative prediction gets 51.31% results and recall obtained from True positive is 51.74% whereas to true negative is 51.47%.

Keywords: *Decision Tree, K-NN, Naive Bayes, Online Lectures, Twitter, Text Mining*

1. Pendahuluan

Kehadiran pandemi *Coronavirus Diseases (Covid-19)*, menyebabkan krisis kesehatan diseluruh dunia. Pada tanggal 9 Maret 2020 *World Health Organization (WHO)* mendeklarasikan bahwa virus covid-19 sebagai

pandemic dan telah menyebar secara luas di berbagai belahan dunia (STPCOVID-19, 2021). DiIndonesia untuk mengurangi penyebaran virus melalui kontak fisik Presiden Joko widodo mengeluarkan kebijakan Pembatasan Sosial Berskala Besar (PSBB). Kebijakan PSBB ini terlampir dalam peraturan Pemerintah No.21 Tahun 2020 tentang Pembatasan social Berskala Besar (PP PSBB) dalam rangka akselerasi penanganan Covid-19 diIndonesia (BPKRI, 2020).

Penerapan PSBB mengakibatkan akibat dan perubahan besar dalam banyak bidang, salah satu yang terdampak dalam penerapan PSBB yakni pada pembatasan aktivitas pembelajaran di sekolah tingkatan dasar, hingga keperguruan tinggi. Sebagai tanggapan diberlakukannya PSBB, pemerintah menghimbau penutupan sekolah selama pandemic Covid-19 diIndonesia. Kementerian Pendidikan dan Kebudayaan (Kemendikbud) mengeluarkan surat edaran Nomor 4 tahun 2020 tentang pelaksanaan kebijakan pendidikan pada masa darurat penyebaran Covid-19 menyatakan bahwa proses belajar dilakukan dirumah melalui pembelajaran daring atau jarak jauh guna memberikan pembelajaran yang lebih bermakna, proses pembelajaran dilaksanakan lewat pembelajaran dalam jaringan (daring) (Jayul & Irwanto, 2020). Kebijakan ini mewajibkan penggunaan jaringan internet melalui *smartphone*, *gadget*, ataupun komputer, sebagai media pembelajaran (Samsir et al., 2021). Aktivitas pembelajaran daring dilakukan agar siswa dapat melaksanakan kegiatan belajar tanpa terhambat akses dan dapat mengikuti suatu pembelajaran atau mata kuliah tertentu tanpa harus tatap muka secara langsung dan pembelajaran daring ini memanfaatkan platform seperti zoom, Google Classroom, Google Meet dan sebagainya sebagai media pembelajaran (Damanik et al., 2021).

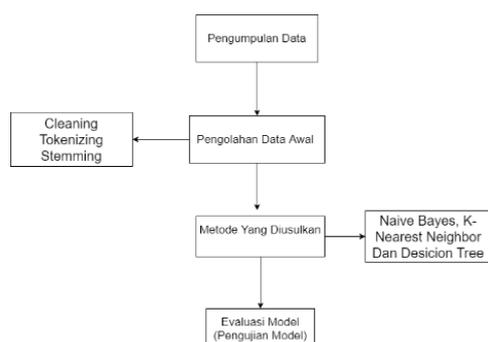
Pembelajaran daring yang disebut sebagai strategi kemudian menjadi kontroversi (Samsir et al., 2021). Merujuk pada luasnya daerah diIndonesia, pembelajaran daring serta merta dapat dilaksanakan dengan mudah, terutama untuk wilayah peserta didik

yang belum terjangkau listrik serta akses internet (Sadikin & Hamidah, 2020). Serta perbedaan infrastruktur, kualitas koneksi internet hingga perangkat yang digunakan sebagai media pembelajaran menjadi hambatan bagi peserta didik untuk melaksanakan pembelajaran secara daring. Kurangnya adaptasi serta perubahan secara mendadak pada skala besar menyebabkan berbagai respon masyarakat (Samsir et al., 2021). Berbagai opini dan komentar dilontarkan masyarakat dilaman berbagai media social, pendapat dan luapan emosi terkait penerapan belajar secara daring.

Salah satunya twitter. Berdasarkan permasalahan diatas penelitian ini bertujuan untuk menganalisis opini publik terhadap implementasi kuliah daring di Indonesia. Dalam penelitian ini penambangan data difokuskan ke bagaimana kuliah daring di Indonesia diterima oleh masyarakat khususnya mahasiswa. Adapun usulan untuk menanggulangi permasalahan pada penelitian kali ini adalah menggunakan metode Naïve Bayes sebagai metode dalam pengolahan text mining terhadap analisis sentimen penerapan perkuliahan daring, dan metode *K-NN* serta *Decision Tree* sebagai metode pembanding dalam penelitian ini. Usulan metode *Naïve Bayes*, *K-NN* dan *Decision Tree* ini didasarkan pada beberapa penelitian terdahulu yang telah membuktikan bahwa metode ini memberikan hasil yang baik dalam pengolahan analisis sentiment dengan objek yang similar. Adapun salah satu penelitian yang dilakukan oleh (Puspita & Widodo, 2021) yang menganalisis Sentimen Pengguna Layanan BPJS menggunakan tiga algoritma, dalam penelitian ini didapatkan hasil algortima *K-NN* mendapatkan nilai akurasi 95,58% dan *class precision negative* sebesar 45,00%, *positive* 0,00%, dan *netral* 96,83%. Pada algortima *Decision Tree* mendapatkan nilai akurasi 96,13% dengan *class precision negative* 55,00%, *positive* 0,00%, dan *netral* 97,28%. Dan *Naïve Bayes* akurasi sebesar 89,14% dengan *class precision pred. negative* 16,67%, *pred. positive* 1,64%, dan *pred. neutral* adalah 98,40%.

2. Metode Penelitian

Penelitian ini terdiri dari beberapa tahapan yang pertama pengumpulan data *tweet* ditwitter, yang kedua praproses data, yang ketiga pengklasifikasi algoritma dan klasifikasi sentimen, dan yang ke empat evaluasi atau pengujian model.



Gambar 1. Alur Penelitian

Tahap Pengumpulan Data

Pada penelitian ini data yang digunakan hanya tweet berbahasa Indonesia, dan tweet yang mengandung opini masyarakat terhadap kuliah daring. Pencarian data dilakukan dengan menggunakan kata kunci “kuliah daring”, tweet yang diambil hanya berupa kalimat dan tidak mengandung gambar. Pengambilan data memanfaatkan Bahasa pemrograman Rapid Miner 9.10 dan menggunakan *Application Program Interface (API)*. API digunakan untuk mengambil data tweet dari server twitter. Jumlah data yang didapat sebanyak 500 tweet dari limit yang diajukan. Pengambilan jumlah data dibatasi agar proses selanjutnya dapat berjalan dengan baik, karena jika terlalu banyak data akan memperlambat proses data.

Pengolahan Data Awal

Pada tahap awal dilakukan preprocessing pada data untuk selanjutnya dilakukan klasifikasi terhadap data-data yang dikumpulkan menggunakan algoritma yang dipilih, adapun tahapan-tahapan analisisnya sebagai berikut:

Preprocessing Data

Praprocess dilakukan untuk menghindari data yang belum terstruktur dan untuk menghasilkan data bersih dan terstruktur yang memudahkan untuk proses klasifikasi dan memberikan hasil yang lebih akurat (Musfiroh et al., 2021). Data yang dimasukkan pada tahap ini berupa data tweet yang mentah. Adapun tahapan praproses meliputi:

a. Tahap Cleaning, adalah tahap pertama praprocessing yang dilakukan untuk membersihkan suatu noise pada data.

Tahap ini terdiri dari beberapa langkah yaitu:

- 1) *Select Attribute* untuk memisahkan kolom yang hanya digunakan, dalam penelitian ini hanya kolom text saja yang dipilih.

- 2) *Remove duplicates*, untuk menghilangkan tweet yang berduplikat atau spam tweet.

- 3) *Replace Missing Value* untuk mengetahui data yang hilang, dan menghilangkan data yang tidak bernilai.

- 4) Subprocess berfungsi sebagai wadah *replace* yang berguna untuk menghapus tanda baca yang tidak penting seperti @, #, Rt, URL, dan lain sebagainya.

Setelah proses cleaning selesai dilakukan, maka akan diperoleh data yang sudah bersih dan dapat dilakukan proses selanjutnya yaitu analisis sentiment, proses analisis sentiment dilakukan untuk mengetahui sentiment yang terkandung dalam data tweet, dalam penelitian ini hanya sentiment positif dan negative yang digunakan dalam analisis.

b. Tahap Tokenizing adalah tahap pemotongan atau pembagian string menjadi urutan token berdasarkan tiap kata yang menyusunnya, atau juga disebut proses pemecah sekumpulan kalimat menjadi satuan kata (Rezki et al., 2020).

c. Tahapan Stemming, Stemming merupakan tahap untuk mencari root kata dari hasil filtering. Atau bisa disebut proses pengubah kata ke bentuk dasar (stem) (Rezki et al., 2020).

Setelah proses tokenizing selesai dilakukan proses selanjutnya yaitu wordcloud, proses wordcloud ini berfungsi untuk mengetahui kata apa saja yang sering dibicarakan atau sering muncul dalam data tweet dalam bentuk wordcloud.

Klasifikasi Naïve Bayes

Naïve Bayes Classifier adalah algoritma Bayesian Learning yang paling cepat dan sederhana. Algoritma klasifikasi *Naïve Bayes* cocok untuk diterapkan pada data dengan jumlah yang banyak dan mampu menangani data yang kosong (missing value) serta menggunakan perhitungan probabilitas dan statistic yang memprediksi peluang dimasa depan berdasarkan pengalaman sebelumnya (Hilda, 2018).

Klasifikasi K-Nearest Neighbor (KNN)

K-Nearest Neighbor ini termasuk ke dalam grup *instance-based, learning*. Metode KNearest Neighbor merupakan teknik *lazy, learning* (Asril et al., 2019). Dimana hasil klasifikasi metode ini berdasarkan kedekatan jarak data.

Klasifikasi Decision Tree

Decision tree merupakan suatu metode klasifikasi dimana variabel dependen yang bertipe kategorik maka menghasilkan pohon klasifikasi (*classification trees*). dimana node paling atas disebut root dan cabangnya berisi nilai dari atribut. (Salman, 2020).

Pengujian Model

Pada tahap pengujian menggunakan Teknik *cross validaton*, data dibagi menjadi dua dimana training set digunakan untuk melatih model, sedangkan testing set digunakan untuk mengevaluasi performa dari model. Pada Teknik *cross validation* dilakukan perulangan untuk menghindari overlapping pada data uji. Yang kemudian diproses dalam bentuk tabel *confusion matrix* (Musfiroh et al., 2021).

Tabel 1. *Confusion Matrix*

		Prediksi	
		Positif	Negatif
Aktual	Positif	TP (True Positif)	FN (False Negatif)
	Negatif	FP (False Positif)	TN (True Negatif)

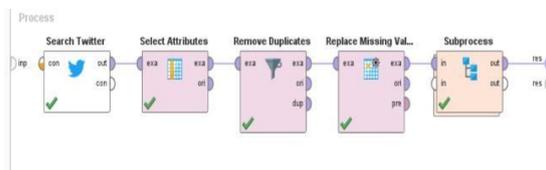
Sumber: (Musfiroh et al., 2021)

3. Hasil dan Pembahasan

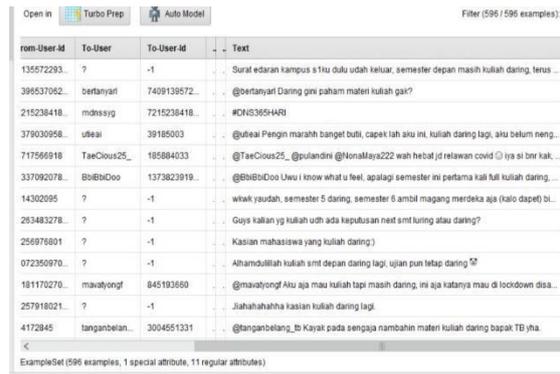
Penerapan metode Naïve Bayes, K-NN dan Decision Tree bertujuan untuk mengklasifikasi sentimen terhadap kuliah daring dan untuk mengetahui perbandingan tingkat akurasi dari dari ketiga algoritma dalam menganalisis sentimen komentar masyarakat pada media sosial Twitter.

3.1. Hasil Pengumpulan Data

Proses pengambilan data dengan menggunakan Rapid Miner seperti gambar 2.



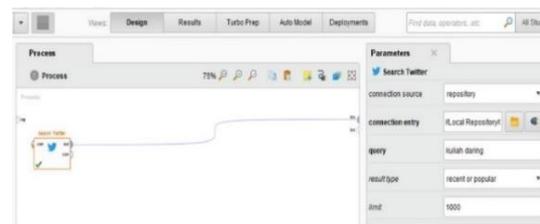
Gambar 2. Proses *Crawling*



Gambar 3. Hasil *Crawling Tweet*

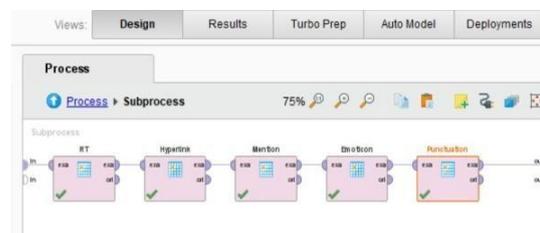
3.2. Hasil Preprocessing

Data tweet yang diperoleh kemudian dilakukan preprocessing data untuk memperoleh data yang lebih terstruktur untuk memudahkan proses klasifikasi sentimen. Pada saat preprocessing data terdiri dari beberapa tahapan, Terdapat 3 tahap dalam proses preprocessing untuk memperoleh hasil yang maksimal, yaitu tahap cleaning, tahap tokenizing dan tahap stemming. Adapun tahapan cleaning meliputi



Gambar 4. Tahapan *Cleaning*

Terdapat 6 *Replace* yang dipilih dalam operator subprocess, dalam masing-masing *Replace* terdapat proses untuk menghilangkan @, RT, #, *Hyperlink*, *Function* dan *Emoticon*.



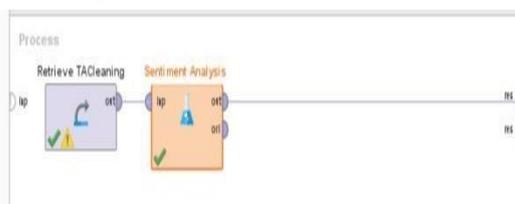
Gambar 5. Proses *Replace*

Setelah proses cleaning selesai, maka didapat data yang sudah bersih dan dapat digunakan pada proses selanjutnya

Tabel 2. Hasil *Crawling* dan *Cleaning*

<i>Data Crawling</i>	<i>Data Cleaning</i>
mahasiswa sekarang sukanya daring.. kuliah aja daring, besok kalau demo pasti juga maunya daring.. gimana kagak, wisuda nya aja daring 😄😄😄	mahasiswa sekarang sukanya daring kuliah aja daring besok kalau demo pasti juga maunya daring gimana kagak wisuda nya aja daring
di up agk siangan menuju sore, krna pasti ada yg masih daring,kuliah dll bye mo tidur dulu buset ngntuk bet 😞 https://t.co/pHTO4CBq3n https://t.co/dU4QVOmF4f	di up agk siangan menuju sore krna pasti ada yg masih daringkuliah dll bye mo tidur dulu buset ngntuk bet
Kebanyakan main games ML Efek kuliah daring https://t.co/0uNKxlV8ZI	Kebanyakan main games ML Efek kuliah daring
@bertanyarl Kayaknya lo lagi stress nder, gue kuliah daring apalagi klo kuis /ujian gitu trs klo diajak ngobrol sama org rumah juga kadang otomatis mereka gue bentak / ngomongnya ngegas, soalnya bikin ga konsen.	Kayaknya lo lagi stress nder gue kuliah daring apalagi klo kuis ujian gitu trs klo diajak ngobrol sama org rumah juga kadang otomatis mereka gue bentak ngomongnya ngegas soalnya bikin ga konsen

selanjutnya yaitu analisis sentimen terhadap data yang sudah bersih.



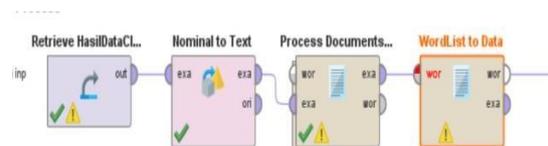
Gambar 6. Proses Analisis Sentimen

Tabel 3. Hasil Analisis Sentimen

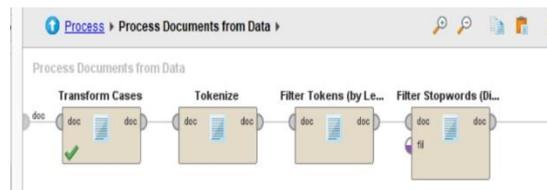
<i>Data Cleaning</i>	<i>Polarity</i>
mahasiswa sekarang sukanya daring kuliah aja daring besok kalau demo pasti juga maunya daring gimana kagak wisuda nya aja daring	P
di up agk siangan menuju sore krna pasti ada yg masih daringkuliah dll bye mo tidur dulu buset ngntuk bet	N
Kebanyakan main games ML Efek kuliah daring	P
Kayaknya lo lagi stress nder gue kuliah daring apalagi klo kuis ujian gitu trs klo diajak ngobrol sama org rumah juga kadang otomatis mereka	N
gue bentak ngomongnya ngegas soalnya bikin ga konsen	

3.3. Hasil *Tokenizing*

Pada proses tokenizing dilakukan pemotongan atau pemisahan setiap kata dalam teks yang disebut sebagai token. Dalam proses ini terdapat beberapa proses meliputi yang pertama Penambahan atribut Nominal to text, atribut ini berfungsi sebagai data nominal menjadi text, yang kedua Penambahan Proses Documen From Data, atribut ini berfungsi sebagai wadah atau tempat menyimpan proses tokenizing dan yang ketiga Penambahan proses Wordlist, berfungsi untuk mengkonveksi daftar kata menjadi penggalan kata.

Gambar 7 Proses *Wordlist Tokenizing*

Adapun 4 operator yang dijalankan dalam proses tokenizing, yang pertama Transform cases, elemen ini berfungsi untuk mengubah data menjadi huruf kecil, yang kedua tokenize untuk tokenisasi atau pemecahan string kalimat menjadi kata, yang ketiga filter tokens berfungsi untuk memfilter kata minimal dan maksimal karakter, dan yang keempat ada filter Stopword yang berfungsi untuk pengambilan kata-kata penting hasil token (membuang kata kurang penting dan menyimpan kata penting).



Gambar 8. Proses Tokenizing

Berikut contoh hasil dari proses tokenizing dapat dilihat pada table 4. Hasil Tokenizing.

Tabel 4. Hasil Tokenizing

Sebelum Proses	Sesudah Proses
mahasiswa sekarang sukanya daring kuliah aja daring besok kalau demo pasti juga maunya daring gimana kagak wisudanya aja daring	'mahasiswa' 'sekarang' 'sukanya' 'daring' 'kuliah' 'daring' 'daring' 'besok' 'kalau' 'demo' 'pasti' 'juga' 'maunya' 'daring' 'gimana' 'kagak' 'wisuda' 'aja' 'daring'
di up agk siangan menuju sore krna pasti ada yg masih daringkuliah dll bye mo tidur dulu buset ngntuk bet	'Agak' 'siangan' 'menuju' 'sore' 'krna' 'pasti' 'ada' 'masih' 'daring' 'kuliah' 'dll' 'bye' 'tidur' 'dulu' 'buset' 'ngntuk' 'banget'
Kebanyakan main games ML Efek kuliah daring	'Kebanyakan' 'main' 'games' 'Efek' 'kuliah' 'daring'
Kayaknya lo lagi stress nder gue kuliah daring apalagi klo kuis ujian gitu trs klo diajak ngobrol sama orang rumah juga kadang otomatis mereka gue bentak ngomongnya ngegas soalnya bikin engga konsen	'Kayaknya' 'lagi' 'stress' 'nder' 'gue' 'kuliah' 'daring' 'apa' 'lagi' 'klo' 'kuis' 'ujian' 'gitu' 'trs' 'diajak' 'ngobrol' 'sama' 'orang' 'rumah' 'juga' 'kadang' 'otomatis' 'mereka' 'bentak' 'ngomong' 'ngegas' 'soal' 'nya' 'bikin' 'engga' 'konsen'

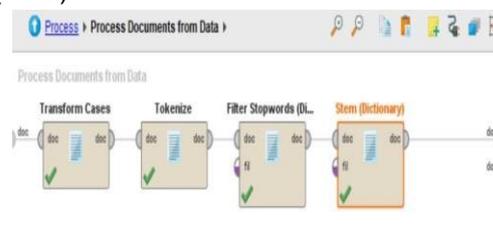
Berdasarkan Gambar 9. menunjukkan perhitungan frekuensi data terdapat beberapa kata kunci, pada data tweet didapatkan kata 'kuliah' disebutkan sebanyak kali 433, disusul dengan kata "daring" sebanyak 422 dan seterusnya.

word	in docu... ↓	total
kuliah	388	433
daring	384	422
dan	81	102
ya	74	88
yang	74	106
aku	73	110
mau	68	72
ini	65	74

Gambar 9. Wordlist Tokenizing

3.4. Hasil Stemming

Pada proses sebelumnya sudah menghasilkan token yang sudah bersih kemudian dilakukan proses pengembalian kata yang sama kedalam bentuk dasarnya (stem)



Gambar 10. Proses Stemming

Tabel 5. Hasil Stemming

Sebelum Proses	Sesudah Proses
Gak, gk jg krn agk tp cm krna bgt, bet yg	tidak juga karena agak tapi Cuma Karena Banget Yang

3.5. Hasil Klasifikasi Model

Dalam tahapan ini data kemudian diklasifikasi menggunakan algoritma yang sudah dipilih dan disentimen kedalam dua kelas (positif dan negative).

kemunculan, sedangkan "Kuliah" memiliki 433 kemunculan dengan masing-masing presentase sebesar 100% dibanding kata-kata yang lain.

3.7. Hasil Pengujian Model

Setelah proses preprocessing, selanjutnya akan dilakukan proses pengujian model melalui tahap validation, pada tahap ini dilakukan pengukuran atau perhitungan tingkat akurasi dari tiap algoritma yang meliputi *Accuracy*, *Precision* dan *Recall*. Dan akan disajikan dalam bentuk table confusion matrix.

Pengujian Naïve Bayes

	true P	true N	class precision
pred. P	214	0	100.00%
pred. N	125	339	73.06%
class recall	63.13%	100.00%	

Gambar 19. Hasil *performance* algoritma naïve bayes

1. Accuracy

$$\begin{aligned} & TP + TN \\ \text{accuracy} &= \frac{TP + TN}{TP + FP + FN + TN} \times 100\% \\ &= \frac{214 + 0 + 125 + 339}{214 + 339} \times 100\% \\ &= \frac{553}{678} \times 100\% \\ &= \frac{55300}{678} = 81,57\% \end{aligned}$$

2. Precision

$$\begin{aligned} & TP \\ \text{precision Positif} &= \frac{TP}{TP + FP} \times 100\% \\ &= \frac{214 + 0}{214} \times 100\% \\ &= \frac{214}{214} = 100\% \\ & TN \\ \text{precision Negatif} &= \frac{TN}{TN + FN} \times 100\% \\ &= \frac{339 + 125}{33900} \times 100\% \\ &= \frac{464}{33900} = 73,06\% \end{aligned}$$

3. Recall

TP

$$\text{recall Positif} = \frac{TP}{TP + FN} \times 100\%$$

$$\begin{aligned} & \frac{214}{214 + 125} \times 100\% \\ &= \frac{214}{21400} \times 100\% \\ &= \frac{339}{21400} = 63,13\% \end{aligned}$$

TN

$$\text{Recall Negatif} = \frac{TN}{TN + FP} \times 100\%$$

$$\begin{aligned} & \frac{339}{339 + 0} \times 100\% \\ &= \frac{33900}{33900} \times 100\% \\ &= \frac{339}{339} = 100\% \end{aligned}$$

Selanjutnya hasil dari nilai akurasi akan menghasilkan *Area Under Curve (AUC)*.



Gambar 20. Grafik *Area Under Curve (AUC)* Metode Naïve Bayes

Area Under Curve (AUC) yang dihasilkan adalah 1000 dimana keakuratan algoritma naïve bayes masuk kedalam kelompok *Excellent Classification*.

Pengujian K-NN

Berikut merupakan hasil dari *confusion matrix* menggunakan algoritma *K-Nearest Neighbor* berdasarkan percobaan nilai K-5, ditunjukkan pada gambar 21.

	true P	true N	class precision
pred. P	211	129	62.06%
pred. N	128	210	62.13%
class recall	62.24%	61.95%	

Gambar 21. *Performance* Algoritma *K-Nearest Neighbor (K-NN)*

1. Accuracy

$$\frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

$$= \frac{211 + 210}{211 + 129 + 128 + 210} \times 100\%$$

$$= \frac{42100}{678} = 62,10\%$$

2. Precision

$$\frac{TP}{TP + FP} \times 100\%$$

$$= \frac{211}{211 + 129} \times 100\%$$

$$= \frac{21100}{340} = 62,06\%$$

TN

$$\frac{TN}{TN + FN} \times 100\%$$

$$= \frac{210}{210 + 128} \times 100\%$$

$$= \frac{21000}{339} = 62,13$$

3. Reccall

$$\frac{TP}{TP + FN} \times 100\%$$

$$= \frac{211}{211 + 128} \times 100\%$$

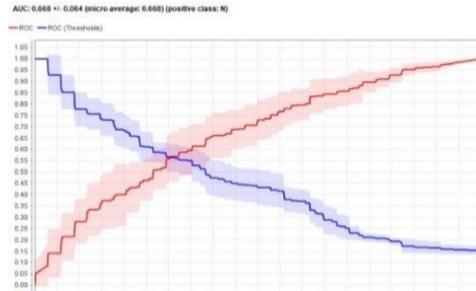
$$= \frac{21100}{339} = 62,24\%$$

TN

$$\frac{TN}{TN + FP} \times 100\%$$

$$= \frac{210}{210 + 129} \times 100\%$$

$$= \frac{21000}{339} = 61,95\%$$



Gambar 22. Grafik Area Under Curve (AUC) Metode K-Nearest Neighbour (K-NN)

Under Curve (AUC) yang dihasilkan adalah 0.668 dimana keakuratan algoritma KNN masuk kedalam kelompok *Poor Classification*

3.7.3. Pengujian Decision Tree

accuracy: 51.18% +/- 1.00% (micro average: 51.18%)

	true P	true N	class precision
pred. P	173	165	51.18%
pred. N	166	174	51.18%
class recall	51.03%	51.33%	

Gambar 23. Performance Decision Tree

1. Accuracy

$$\frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

$$= \frac{178 + 175}{178 + 165 + 166 + 175} \times 100\%$$

$$= \frac{35300}{684} = 51,60\%$$

2. Precision

$$\frac{TP}{TP + FP} \times 100\%$$

$$= \frac{178}{178 + 165} \times 100\%$$

$$= \frac{17800}{343} = 51,89\%$$

TN

$$\frac{TN}{TN + FN} \times 100\%$$

$$= \frac{175}{175 + 166} \times 100\%$$

$$\frac{17500}{341} = 51,31\%$$

3. *Recall*

TP

$$\text{recall Positif} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100\%$$

$$= \frac{178}{178 + 166} \times 100\%$$

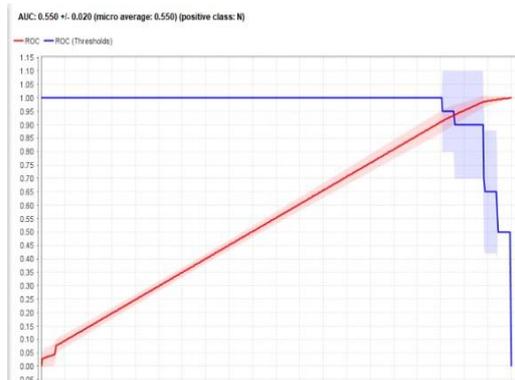
$$= \frac{17800}{344} = 51,74\%$$

TN

$$\text{recall Negatif} = \frac{\text{TN}}{\text{TN} + \text{FP}} \times 100\%$$

$$= \frac{175}{175 + 165} \times 100\%$$

$$= \frac{17500}{340} = 51,47\%$$



Gambar 24. Grafik Area Under Curve (AUC) Metode Decision Tree

Area Under Curve (AUC) yang dihasilkan adalah 0.550 dimana keakuratan algoritma Decision tree masuk kedalam kelompok *Fatluure Classification*

4. Kesimpulan

1. Pada penelitian ini, upaya untuk mengetahui opini publik di media twitter terhadap berlakunya kuliah daring dimasa pandemi Covid-19 oleh pemerintahan di Indonesia dapat menjadi masukan bagi indikator keberhasilan pemerintah. Dengan menggunakan algoritma *Naive Bayes*, *K-NN* dan *Decision Tree* untuk klasifikasi tweet menghasilkan ketiga algoritma bekerja dengan baik. Hasil akurasi menunjukkan algoritma *naive bayes* menghasilkan nilai akurasi sebesar

81,57% pada algoritma *K-Nearest Neighbor* didapatkan akurasi sebesar 62,10%. Dan untuk Algoritma *Decision Tree* hasil akurasi sebesar 51,89%.

2. Dalam penelitian ini, dapat dibuktikan algoritma *Naive Bayes* merupakan algoritma yang memiliki nilai akurasi paling tinggi dibandingkan kedua metode lainnya dengan tingkat akurasi sebesar 81,57%.
3. Dengan uji validasi metode *Naive Bayes Classifier* menggunakan *cross validation* didapatkan rata-rata untuk nilai *Precision* positif 100% dan *Precision* Negatif 73,06% sebesar, *Recall* Positif 63,13% dan *Recall* Negatif sebesar 100%. Sedangkan untuk uji validasi dengan metode *K-NN* didapatkan nilai rata-rata *Precision* positif 62,06% dan *Precision* Negatif 62,13% sebesar, *Recall* Positif 62,24% dan *Recall* Negatif sebesar 61,95%. Dan untuk *Decision Tree* *Precision* Positif 51,89%, *Precision* Negatif 51,31%, *Recall* Positif 51,74% dan *Recall* Negatif sebesar 51,47%.
4. Dari hasil analisis sentiment yang diperoleh terdapat sentiment positif lebih besar dari pada sentiment negative yaitu sebesar 5,57% dan sentiment negative sebesar 94,43% ini menunjukkan bahwa sebagian mahasiswa dapat menerima penerapan kuliah daring dimasa pandemic covid-19 seperti sekarang ini, ini dibuktikan dengan beberapa tweet yang menyebutkan bahwa kuliah daring cukup efektif dimasa pandemic seperti sekarang ini, namun masih banyak juga mahasiswa yang mengeluhkan tentang pembayaran UKT yang dibebankan kepada mahasiswa dimasa kuliah daring tanpa harus merasakan fasilitas atau pembelajaran secara tatap muka dikampus, serta biaya listrik dan kuota yang tinggi karena penerapan daring.

Referensi

- Asril, H., Mustakim, & Kamila, I. (2019). Klasifikasi Dokumen Tugas Akhir Berbasis Text Mining menggunakan Metode *Naive Bayes Classifier* dan *K-Nearest Neighbor*. *Seminar Nasional*

- Teknologi Informasi Komunikasi Dan Industri*, 0(0), 2579–5406. <http://ejournal.uin-suska.ac.id/index.php/SNTIKI/article/view/7995>
- BPKRI. (2020). *Pembatasan Sosial Berskala Besar dalam Rangka Percepatan Penanganan Penyakit Virus Corona 2019 (COVID-19)*. JDIIH BPK RI DATABASE. <https://peraturan.bpk.go.id/Home/Details/135059/pp-no-21-tahun-2020>
- Damanik, A. R., Sumijan, & Nurcahyo, G. W. (2021). *Jurnal Sistim Informasi dan Teknologi Prediksi Tingkat Kepuasan dalam Pembelajaran Daring Menggunakan Algoritma Naïve Bayes*. 3, 88–94. <https://doi.org/10.37034/jsisfotek.v3i3.49>
- Hilda, A. (2018). *PERBANDINGAN METODE DATA MINING SVM DAN NN UNTUK KLASIFIKASI PENYAKIT GINJAL KRONIS*. 14(1), 1–6.
- Jayul, A., & Irwanto, E. (2020). Model Pembelajaran Daring Sebagai Alternatif Proses Kegiatan Belajar Pendidikan Jasmani di Tengah Pandemi Covid-19 Achmad. *Jurnal Pendidikan Kesehatan Rekreasi*, 6(2), 190–199.
- Musfiroh, D., Khaira, U., Eko, P., Utomo, P., & Suratno, T. (2021). *Sentiment Analysis of Online Lectures in Indonesia from Twitter Dataset Using InSet Lexicon Analisis Sentimen terhadap Perkuliahan Daring di Indonesia dari Twitter Dataset Menggunakan InSet Lexicon*. 1(April), 24–33.
- Puspita, R., & Widodo, A. (2021). Perbandingan Metode KNN, Decision Tree, dan Naïve Bayes Terhadap Analisis Sentimen Pengguna Layanan BPJS. *Jurnal Informatika Universitas Pamulang*, 5(4), 646. <https://doi.org/10.32493/informatika.v5i4.7622>
- Rezki, M., Kholifah, D. N., Faisal, M., Priyono, P., & Suryadithia, R. (2020). Analisis Review Pengguna Google Meet dan Zoom Cloud Meeting Menggunakan Algoritma Naïve Bayes. *Jurnal Infortech*, 2(2), 264–270. <https://doi.org/10.31294/infortech.v2i2.9286>
- Sadikin, A., & Hamidah, A. (2020). Pembelajaran Daring di Tengah Wabah Covid-19. *Biodik*, 6(2), 109–119. <https://doi.org/10.22437/bio.v6i2.9759>
- Salman, A. (2020). *Konsep Decision Tree & Random Forest*. Binus University. <https://socs.binus.ac.id/2020/05/26/konsep-decision-tree-random-forest/>
- Samsir, Ambiyar, Verawardina, U., Edi, F., & Watrianthos, R. (2021). Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naïve Bayes. *Jurnal Media Informatika Budidarma*, 5, 157–163. <https://doi.org/10.30865/mib.v5i1.2604>
- STPCOVID-19. (2021). *Apa Yang Dimaksud Dengan Pandemi*. Komite Penanganan Covid-19 Dan Pemulihan Ekonomi Nasional. <https://covid19.go.id/tanya-jawab?search=Apa yang dimaksud dengan pandemi>