IJCSR: The Indonesian Journal of Computer Science Research
Volume 1, Nomor 2 Juli 2022
https://subset.id/index.php/IJCSR

Perbandingan Algoritma Naïve Bayes dan Decision Tree Pada Sentimen Analisis

I Putu Wibina Karsa Gumi[1], Hartatik[2], Andri Syafrianto[3, a)]

[1,2)] *Fakultas Ilmu Komputer*
Universitas Amikom Yogyakarta, Jl. Ringroad Utara, Depok, Condongcatur, Sleman, Yogyakarta, Indonesia
[3)]*Program Studi Informatika*
STMIK EL RAHMA Yogyakarta, Jl. Sisingamangaraja Jl. Karangkajen No.76, Brontokusuman, Kec. Mergangsan, Kota Yogyakarta

Author Emails
[a)] Corresponding author: andrisyafrianto@gmail.com

Abstract. *The development of information technology in recent year is really fast which pushing the usage of internet. People use social media to share their opinion or just as simple as interacting with each other. Sentiment analysis is a branch of Natural Language Processing (NLP) that can filter and categorizing people opinion on social media. This study uses twitter data to compare two algorithm that is Naïve Bayes Classifier and Decision Tree. This study divided the data into two scenarios where the first one with 800 data and the second one with 200 data. Data from each scenario divided again into 70% training data and 30% test data from the total of 1000 data. The result show that Naïve Bayes Classifier have much higher accuracy with 85% compared to Decision Tree with 78% on the second scenario.*

Keywords**:** *Comparation, Naïve Bayes, Decision Tree, Sentiment Analysis*

Abstraksi. *Perkembangan teknologi informasi dalam beberapa tahun terakhir sangat pesat dimana hal tersebut mendorong penggunaan internet dan pertukaran informasi. Beragam sosial media digunakan masyarakat untuk membagi opini mereka atau sekadar berinteraksi dengan orang lain. Analisis Sentimen adalah cabang dari Natural Language Processing (NLP) yang dapat menyaring dan mengkategorikan opini masyarakat pada sosial media. Penelitian ini memanfaatkan data dari twitter untuk membandingakan dua algoritma klasifikasi yaitu Naïve Bayes Classifier dan Decision Tree. Penelitian dilakukan dengan membagi data menjadi dua skenario dimana skenario 1 dengan 800 data dan skenario 2 dengan 200 data. Data masing-masing skenario dibagi menjadi 70% data latih dan 30% data uji dari total 1000 data. Hasil dari percobaan menunjukan bahwa Naïve Bayes memilki akurasi tertinggi dengan 85% dibandingkan Decision Tree dengan 78% pada skenario kedua*

Kata Kunci: *Perbandingan, Naive Bayes, Decision Tree, Sentimen Analisis*

PENDAHULUAN

Pada masa serba digital sekarang setiap individu dapat memberikan opini mereka di beberapa media termasuk media berita dan media sosial. Salah satu media social tersebut adalah twitter. Dengan mengirim tweet atau cuitan yang merupakan pesan singkat pengguna dapat menyampaikan opini dan pandangannya terhadap suatu topik yang sedang hangat maupun topik umum. Penggunaan twitter bukan hanya untuk sebagai media berbagi informasi namun media tersebut juga sering digunakan untuk bersosialisasi antar pengguna dan membagikan keluh kesah, opini serta sentiment mereka terhadap suatu topik maupun isu yang sedang hangat-hangatnya. Opini dan sentiment tersebut tidak

selalu bernilai positif namun juga dapat bernilai negative. Data opini di twitter tersebut dapat dimanfaatkan untuk mengetahui dan menganalisis opini dan sentiment orang-orang terhadap topik dan isu tersebut.

Salah satu cara untuk menganalisis sentimen dari opini tersebut adalah dengan menggunakan Analisis Sentimen melalui Machine Learning. Machine Learning memiliki beberapa teknik yaitu *Supervised* dan *Unsupervised* dimana keduanya dapat digunakan untuk menganalisa sentimen dari opini tersebut[1]. Beberapa metode pada supervised learning diantaranya adalah Naïve Bayes, K-nearest Neighboor, Decision Tree[2], Support Vector Machine, dan Random Forest[3]. Berdasarkan hal tersebut maka penelitian ini akan melakukan perbandingan antara algoritma Naïve Bayes dan Decision Tree pada sentimen analisis dengan bantuan TF-IDF untuk membobotkan kata dalam dokumen. Penilaian untuk membandingkan kedua algoritma tersebut selanjutnya akan menggunakan Confusion Matrix dengan mengukur tingkat akurasi, presisi, dan recall.

TINJAUAN PUSTAKA

Beberapa penelitian mengenai analisis sentiment dengan menggunakan Naïve Bayes pernah dilakukan dalam menganalisa pembelajaran daring di masa pandemi covid-19 dengan hasil bahwa 30% memberikan respon positif , 60% memberikan respon negative dan 1% netral[4], Selain itu Naïve Bayes juga pernah digunakan untuk menganalisa opini masyarakat terhadap vaksin covid-19 dengan menggunakan 3780 tweet dengan hasil 60,3% memberikan tanggapan positif, 34,4% memberikan tanggapan negative, serta 5,4% memberikan tanggapan menentang/kurang baik dengan nilai akurasi 93%[5]. Nilai akurasi Naïve bayespun pernah dibandingkan dengan SVM untuk mengetahui sentiment masyarakat terhadap suatu kampus, hasilnya diketahui bahwa akurasi Naïve Bayes ternyata lebih baik bila dibandingkan dengan SVM dengan akurasi sebesar 73,65%[6]. Decision tree juga pernah dibandingkan dengan KNNdan Naïve Bayes dalam mengetahui sentiment masyarakat terhadap penggunaan layanan BPJS, berdasarkan ketiga diketahui bahwa akurasi tertingggi berada di decision tree dengan nilai akurasi 96,13%, kemudian KNN dengan nilai akurasi sebesar 95,58% dan Naïve Bayes sebesar 89,14%[7]. Akurasi decision tree yang tinggi ini sejalan ketika decision tree dibandingkan dengan SVM untuk menganalisa sentiment masyarakat terhadap komentar aplikasi transportasi online, dengan hasil akurasi yang dihasilkan decision tree sebesar 90,20% dan SVM sebesar 89,90%[8].

Sentiment Analisis

Sentimen analisis adalah sebuah studi yang menganalisa opini, pendapat, sentiment, penilaian, perilaku, sikap, dan emosi masyarakat terhadap suatu entitas yang dapat berupa topik, produk, layanan, individu, organisasi, peristiwa, fenomena dan masalah. Sentimen analisis biasa juga disebut opinion mining, opinion extraction, dan sentiment mining yang berada pada ranah analisis sentiment [1]. Analisis sentiment berfokus terhadap opini masyarakat yang menunjukan sentimen positif atau sentimen negatif. Analisis sentiment berkembang dan masuk menjadi jenis penelitian dalam natural langague processing (NLP). Klasifikasi sentiment dapat dilakukan dengan dua cara [1] supervised dan unsupervised. Keberhasilan teknik tersebut bergantung pada ekstraksi karakteristik untuk mendeteksi sentiment. Teknik supervised yang paling umum digunakan adalah Support Vector Machine (SVM) dan Naïve Bayes Classifier (NB) [1]. Sedangkan unsupervised menggunakan klasifikasi beberapa tipe kata berdasarkan orientasi semantiknya.

Naïve Bayes

Klasifikasi Naïve Bayes adalah teknik pengklasifikasian yang digunakan untuk memprediksi keanggotaan suatu kelas. Pengklasifikasian Naïve Bayes didasari oleh teori bayes yang memiliki klasifikasi yang sama dengan decision tree [3]. Dasar dari teknik Klasifikasi Naïve Bayes adalah dari teorema bayes dengan kemampuan sebanding dengan decision tree dan neural network. Teori dan persamaan Naïve bayes terangkum pada persamaan (1)

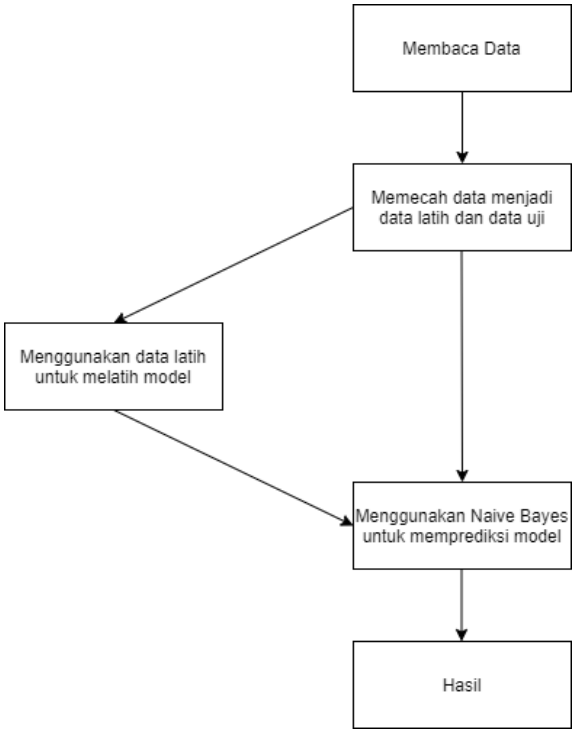
$$P(y|X) = \frac{P(X|y)P(y)}{P(X)} \quad \text{… pers (1)}$$

Dimana :

P(y|X) : adalah Probrabilitas y berdasarkan X

P(X|y) : adalah Probrabilitas X
P(y) : adalah Probabilitas dari y
P(X) : adalah Probabilitas dari X
y : adalah Hipotesa data X
X : adalah Data yang belum diketahui

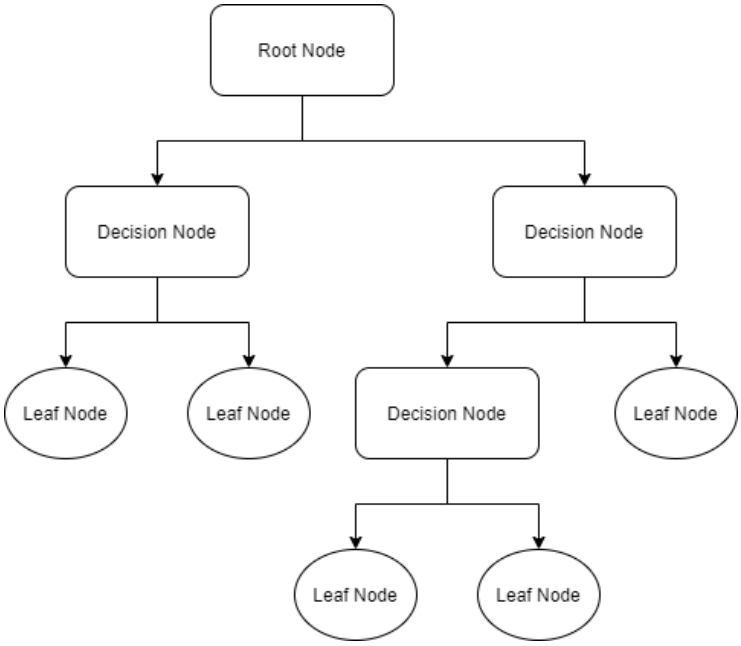
Sedangkan alur dari Naïve Bayes terangkum pada Gambar 1.



GAMBAR 1. Alur Naïve Bayes

Decision Tree

Dalam data mining, algortima klasifikasi dapat mengendalikan data dengan jumlah besar salah satu teknik yang digunakan untuk membuat klasifikasi adalah Decision Tree. Decision Tree adalah salah satu metode yang banyak digunakan dalam berbagai hal misalnya machine lelarning, image processing dan patern identification [5]. Decision Tree adalah sebuah diagram yang dimulai dengan satu node dan kemudian node tersebut memiliki cabang untuk setiap pilihannya yang setiap cabang tersebut akan membuat cabang baru. Alur dari decision tree terangkum pada Gambar 2



Gambar 2. Alur Decision Tree

METODE PENELITIAN

Tahapan penelitian yang akan digunakan dalam penelitian ini adalah:

1. Pengumpulan Data tweet dari twitter menggunakan tweepy
2. Preprocessing data
3. Pembagian Data menjadi data latih dan data uji
4. Pengujian model klasifikasi Naïve Bayes dan Decision Tree

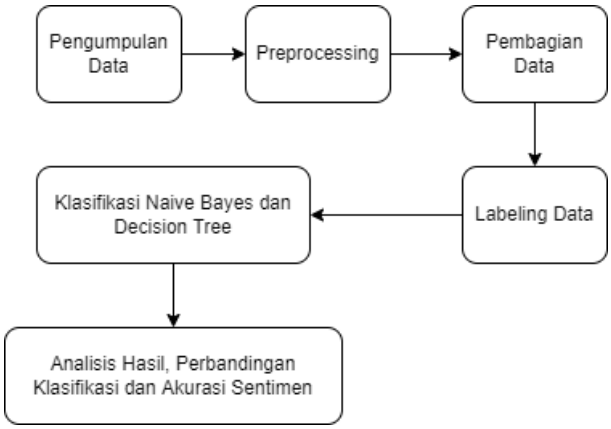
Dengan memanfaatkan data tweet dari twitter makan akan digunakan API twitter untuk mendapatkan dataset yang keseluruhan dataset akan dibagi kedalam dua scenario yaitu dataset kecil dan dataset besar. Setelah mendapatkan dataset proses selanjutnya adalah preprocessing untuk membersihkan dan membuat data lebih terstruktur yang nantinya akan diberikan label, pembobotan, dan ekstraksi fitur dengan bentuan TF-IDF. Proses klasifikasi akan dibagi menjadi 4 dengan masing-masing klasifikasi diberikan 2 dataset, pengukuran nilai akurasi akan menggunakan Confusion Matrix untuk membandingkan efektifitas kedua algoritma tersebut. alur penelitian yang akan dilakukan terangkum pada Gambar 3.

Pengambilan Data

Dataset diambil dari media social Twitter (https://twitter.com) dengan menggunakan API dan program open source Tweepy untuk mengambil data tweet sebanyak 1000 data tweet yang akan terbagi menjadi 2 skenario dimana skenario 1 akan menggunakan 200 data tweet dan scenario 2 akan menggunakan 800 data tweet yang terbagi menjadi data latih dan data uji. Informasi dari tweet yang diambil adalah:

1. Informasi tanggal, bulan, tahun dari tweet
2. Username
3. Isi Tweet

Pengambilan Informasi waktu dan username adalah untuk mengantisipasi adanya tweet yang terduplikasi akibat fitur twitter "retweet" yang memungkinkan pengguna untuk merespon sebuah tweet namun dengan memposting ulang tweet tersebut sebagai "quoted tweet".



GAMBAR 3. Alur penelitian

Preprocessing dan Labeling

Setelah dataset terkumpul dan terbagi maka data tersebut harus dibersihkan agar mempermudah dan membuat data tersebut dapat diolah menjadi informasi baru. Tahapan preprocessing meliputi[9]:

1. Cleaning
2. Case Folding
3. Stemming
4. Filtering (Stopword Removal)
5. Tokenizing

GAMBAR 4. Alur Preprocessing

Labeling dilakukan secara manual oleh peneliti dan Ekstraksi Fitur akan dilakukan dengan menggunakan *metode Term Frequency – Inverse Document Frequency* (TF-IDF) yang bertujuan untuk menilai seberapa penting setiap kata yang muncul dalam data dokumen dan mengubah kata menjadi numerik[10]. Nilai tinggi dan frekuensi kemunculan pada kata dan dokumen akan menunjukan bahwa kata tersebut penting dan umum[11]. Bobot hubungan antara kata dan dokumen akan menunjukan nilai tinggi jika kemunculan kata tersebut tinggi dalam dokumen.

Perancangan Sistem

Dengan menggunakan data yang telah ditarik dari twitter, peneliti akan menggunakan metode TF-IDF sebagai pembobot dan ekstraksi fitur untuk setiap dokumen tweet dan Naïve Bayes dan Decision Tree sebagai pengklasifikaasi dan membandingkan kedua scenario tersebut dengan Confusion Matrix untuk mengukur akurasi dari model klasifikasi.

Analisis Hasil

Analisa dilakukan dengan memanfaatkan Confussion Matrix yang membandingkan dan menghitung nilai akurasi, presisi, dan recall pada setiap metode dengan tujuan untuk mengetahui apakah akurasi, presisi, dan recall dapat diperngaruhi dengan banyaknya jumlah data yang digunakan.

HASIL DAN PEMBAHASAN

Data yang digunakan dalam penelitian ini adalah kumpulan data tweet dari twitter yang didapatkan dengan crawling data dengan API yang disediakan oleh twitter. Data yang diperoleh berupa kumpulan tweet yang disimpan kedalam dokumen dengan format .csv. Untuk memperoleh izin untuk menggunakan API tersebut harus melalui twitter developer mode yang dapat menggunakan source code seperti Gambar 5.

```
access_token="752781860044890113-Ic93QDrhlkDASyBw3ADCies8qamp9au"
access_token_secret="kLPSk64kgd6Z1wWgQXAn5QuWitG3FwGFIDDiMXeFHhXd5"
consumer_key ="HA1zfKmOnSViZF1N2zw6zq49K"
consumer_secret="ItcxVUqmL2LS8hzxG1xst9gujbwUye2JaLOaVqD9zJWhsULLNA"
```

GAMBAR 5. Access Token API

Script tersebut untuk mendapatkan akses untuk API twitter dan kemudian dilanjutkan dengan mengambil data dengan tweepy, source code lengkapnya terangkum pada Gambar 6.

```
import csv
import string
import tweepy
from tweepy import OAuthHandler

access_token="752781860044890113-Ic93QDrhlkDASyBw3ADCies8qamp9au"
access_token_secret="kLPSk64kgd6Z1wWgQXAn5QuWitG3FwGFIDDiMXeFHhXd5"
consumer_key ="HA1zfKmOnSViZF1N2zw6zq49K"
consumer_secret="ItcxVUqmL2LS8hzxG1xst9gujbwUye2JaLOaVqD9zJWhsULLNA"

auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
auth.set_access_token(access_token, access_token_secret)
api = tweepy.API(auth, wait_on_rate_limit=True)

csvFile = open('dataset2.csv', 'a', encoding='utf-8')
csvWriter = csv.writer(csvFile)

search = "vaksin pemerintah covid"
newsearch = search + " -filter:retweets"

for tweet in tweepy.Cursor(api.search,q=newsearch,
tweet_mode="extended", lang="id").items(2000):
    # 2000 item minimalisir tweet duplikat
    print (tweet.full_text)
    csvWriter.writerow([tweet.full_text])
```

GAMBAR 6. Crawling Data Tweet

Selanjutnya akan dilakukan tahapan cleaning yang akan digunakan untuk menghilangkan tanda baca, angka, dan beberapa karakter unik seperti @ yang biasa digunakan sebagai penanda username, # (hashtag), url, dan karakter html. Proses ini dilakukan karena karakter tersebut tidak mempunyai nilai yang signifikan dan tidak memiliki pengaruh pada sentimen pada twitter. Oleh sebab itu karakter tersebut akan dihilangkan untuk membuat dokumen

menjadi bersih dan lebih mudah untuk dibaca dan diproses selanjutnya. Contoh cleaning yang akan dilakukan terangkum pada Tabel 1

TABEL 1. Cleaning

Normal	Cleaning
Vaksin dari pemerintah siap didistibusikan ya #ayovaksinasi	Vaksin dari pemerintah siap didistribusikan ya ayovaksinasi
Pemerintah targetkan vaksinasi hingga akhir 2021	Pemerintah targetkan vaksinasi hingga akhir 2021

Gambar 7 adalah script untuk melakukan cleaning

```
def remove_punct(tweetBersih):
      tweetBersih = re.sub("(@[A-Za-z0-9]+)|([^0-
      9A-Za-     z \t])|(\w+:\/\/\S+)"," ",
      tweetBersih)
      tweetBersih = re.sub('RT[\s]+', '',
      tweetBersih)
      return tweetBersih

tweet_df['text']=tweet_df['text'].apply(lambda x:
remove_punct(x))
```

GAMBAR 7. Cleaning data

Selanjutnya akan dilakukan Case folding yang akan mengubah setiap huruf menjadi lowercase. Contoh proses perubahan dari case folding terangkum pada Tabel 2.

TABEL 2. Case Folding

Normal	Case Folded
Karena telah terjamin aman, bermutu, dan berkhasiat, pemerintah mendorong masyarakat untuk menyegerakan vaksinasi	karena telah terjamin aman, bermutu, dan berkhasiat, pemerintah mendorong masyarakat untuk menyegerakan vaksinasi
Mereka salah satu usia Produktif yang harus divaksin. Umur 12 sampai 18 tahun anak Indonesia wajib diberikan vaksin.	mereka salah satu usia produktif yang harus divaksin. umur 12 sampai 18 tahun anak indonesia wajib diberikan vaksin.

Gambar 8 adalah script untuk melakukan case folding:

(script untuk case folding sudah tergabung dengan stemming dikarenakan library satsrawi secara otomatis melakukan case folding untuk tahap stemming)

GAMBAR 8. Case Folding

Selanjutnya akan dilakukan stemming untuk mengubah kata menjadi kata dasar seperti “-mu”, “-kah”, -“kan”, “-nya”. Contoh perubahan kata pada proses stemming seperti Tabel 3.

TABEL 3. Stemming

Normal	Stemming
Vaksin mampu mencapai herd immunity seseorang supaya mencegah penularan COVID-19.	vaksin mampu capai herd immunity orang supaya cegah tular covid-19
Pemetintah akan mendistribusikan vaksin hingga akhir tahun ini	Perintah akan distribusi vaksin hingga akhir tahun

Gambar 9 adalah script untuk melakukan stemming

```
    def steeming(tweetBersih):
        tweetBersih = stemmer.stem(tweetBersih)
        return tweetBersih
    tweet_df['text']                                =
tweet_df['text'].apply(lambda x: steeming(x))
```

GAMBAR 9. Stemming

Proses stopword removal adalah proses untuk menghilangkan kata yang tidak penting seperti kata penghububg "dan", "di", "ke", dan "atau". Contoh stopword removal terdapat pada Tabel 4.

TABEL 4. Stopword Removal

Normal	Cleaning
Alhamdulillah santri di Indonesia juga beberapa sudah menerima vaksin dari pemerintah. Lekas sehat Indonesiaku.	Alhamdulillah santri Indonesia juga beberapa sudah menerima vaksin pemerintah. Lekas sehat Indonesiaku.
Pemerintah siapkan vaksin dan booster	Pemerintah siapkan vaksin booster

Gambar 10 adalah script untuk melakukan stopword removal :

```
    def remove_stopword(tweetBersih):
        tweetBersih = stopword.remove(tweetBersih)
        return tweetBersih

    tweet_df['text']                                =
tweet_df['text'].apply(lambda x: remove_stopword(x))
```

GAMBAR 10. Stopword Removal

Tokenizing adalah proses untuk memisahkan kata dalam suatu kalimat menjadi potongan yang berdiri sendiri atau tunggal. Contoh tokenizing terdapat pada Tabel 5.

TABEL 5. Tokenizing

Normal	Tokenizing
Kepala BIN menyampaikan masyarakat untuk tidak khawatir	["kepala", "BIN", "menyampaikan", "masyarakat", "untuk", "tidak", "khawatir", "karena", "vaksinasi",

karena vaksinasi yang digunakan saat ini aman, halal dan dijamin	"yang", "digunakan", "saat", "ini", "aman", "halal", "dan", "dijamin"]
Vaksin Sinovac aman jadi tidak perlu takut	["Vaksin", "Sinovac". "aman", "jadi", "tidak", "perlu", "takut"]

Gambar 11 adalah script untuk melakukan tokenizing

```
    def tokenizing(tweetBersih):
        tweetBersih = tweetBersih.split()
        return tweetBersih

    tweet_df['text']                                    =
tweet_df['text'].apply(lambda x: tokenizing(x))
```

GAMBAR 11. Tokenizing

Labeling akan dilakukan secara manual untuk setiap dokumeni, label tersebut pernilai 1 atau 0, 1 jika positif dan 0 jika negatif. Data tersebut adalah data tweet yang sudah dilabeli secara manual dengan nilai positif dan negative. Tabel 6 adalah contoh label data tweet

TABEL 6. Labeling Data

Tweet	Sentimen
Semua jenis vaksin yang disediakan oleh pemerintah dipastikan memiliki persetujuan WHO dan Badan POM	Positif
Kebenaran! Jumlah orang yang meninggal akibat vaksin sangat luar biasa. Pemerintah? Diam saja	Negatif

TF-IDF

Penggunaan TF-IDF adalah untuk ekstraksi fitur dan mengubah data kata menjadi numerik untuk proses analisis sentimen. Kata tersebut akan tergabung menjadi sebuah array yang kemudian akan mudah dibaca oleh computer karena computer hanya bisa membaca angka numerik. Gambar 12 adalah script untuk melakukan TF-IDF.

```
from        sklearn.feature_extraction.text        import
TfidfVectorizer
tfidf = TfidfVectorizer(analyzer='word', use_idf=True)
X_train = tfidf.fit_transform(X_train).toarray()
X_test = tfidf.transform(X_test).toarray()
```

GAMBAR 12. TF-IDF

Analisis Hasil Prediksi Sentimen Dengan Naïve Bayes

Sub-bab ini akan membahas tentang analisa hasil prediksi sentimen "vaksinasi pemerintah" dari twitter dengan Naïve Bayes serta bantuan Confusion Matrix untuk menghitung nilai akurasi, presisi dan recall. Tabel 7 adalah table hasil output setiap scenario

TABEL 7. Hasil Prediksi Naïve Bayes

Skenario	Total Data	Data Latih	Data Uji	Aktual		Hasil
Skenario 1	200	140	60	Positif	Negatif	Akurasi
		Prediksi	Positif	27	11	73%
			Negatif	5	17	
Skenario 2	800	560	240	Positif	Negatif	Akurasi

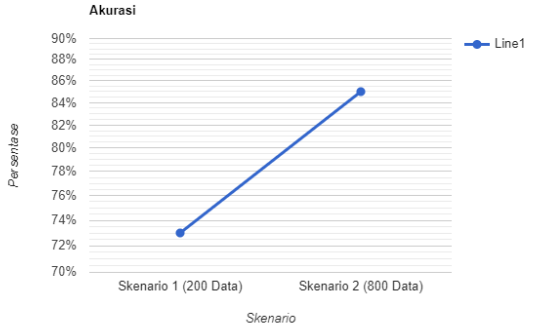
Prediksi	Positif	127	26	85%
	Negatif	10	77	

TABEL 8. Keterangan Tabel Confusion Matrix

	Aktual	
Prediksi	TP (True Positive)	FP (False Positive)
	FN (False Negative)	TN (True Negative)

Akurasi

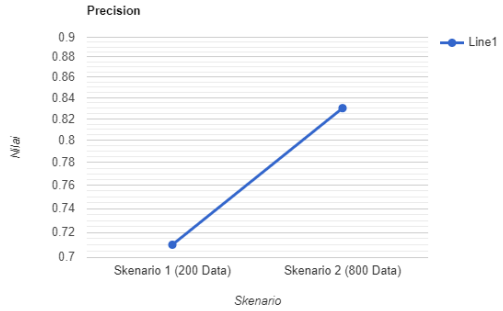
Tingkat akurasi dari hasil analisis dengan algoritma Naïve Bayes dari kedua scenario menunjukan akurasi meningkat dimana pada scenario 1 (200 data) menunjukan akurasi 73% dan scenario 2 (800 data) dengan akurasi tertinggi yakni 85%.



GAMBAR 13. Akurasi Naïve Bayes

Precision

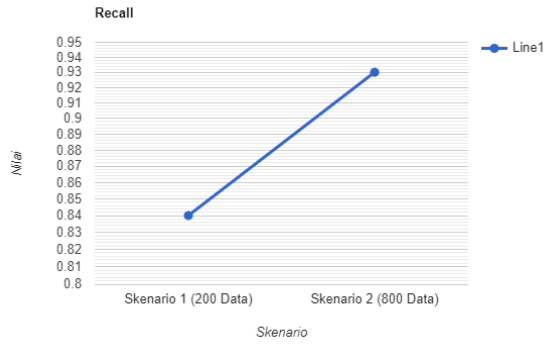
Hasil analisis menunjukan nilai precisioin pada algoritma Naïve Bayes dimana nilai tersebut menunjukan keakuratan klasifikasi dengan 0.71 pada Skenario 1 dan nilai tertinggi 0.83 pada Skenario 2.



GAMBAR 14. Precision Naïve Bayes

Recall

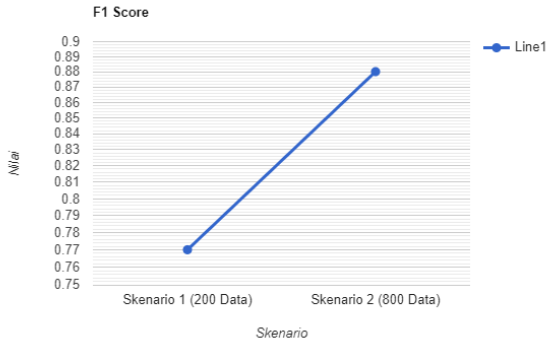
Hasil analisis untuk nilai recall menunjukan apabila semakin tinggi nilai tersebut maka tingkat keberhasilan analisis untuk mengenali kembali suatu dokumen semakin baik. Nilai recall untuk Skenario 1 adalah 0.84 dan untuk scenario 2 memiliki nilai recall tinggi yaitu 0.93.



GAMBAR 15. Recall Naïve Bayes

F1 Score

Hasil analisis Skenario 1 0.77 dan Skenario 2 0.88



GAMBAR 16. F1 Score Naïve Bayes

Hasil Prediksi Sentimen Dengan Decision Tree

Sub-bab ini akan membahas tentang analisa hasil prediksi sentimen "vaksinasi pemerintah" dari twitter dengan Decision Tree serta bantuan Confusion Matrix untuk menghitung nilai akurasi, presisi dan recall. Table 9 adalah table hasil output setiap scenario.

TABEL 9. Hasil Prediksi Decision Tree

Skenario	**Total Data**	**Data Latih**	**Data Uji**	**Aktual**		**Hasil**
Skenario 1	200	140	60	Positif	Negatif	Akurasi
		Prediksi	Positif	27	22	58%
			Negatif	3	8	
Skenario 2	800	560	240	Positif	Negatif	Akurasi
		Prediksi	Positif	110	31	78%
			Negatif	22	77	

TABEL 10. Keterangan Tabel Confusion Matrix

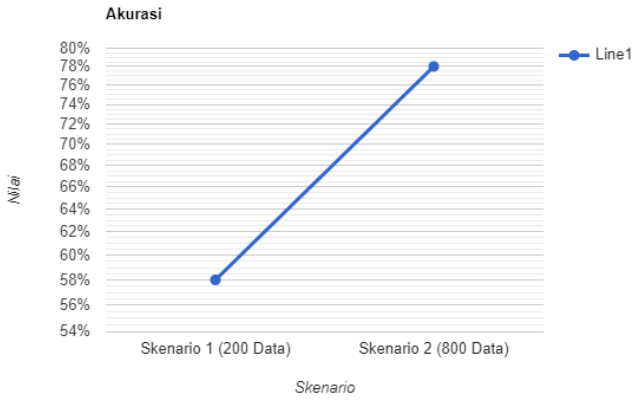
	Aktual	
Prediksi	TP (True Positive)	FP (False Positive)

IJCSR: The Indonesian Journal of Computer Science Research
Volume 1, Nomor 2 Juli 2022
https://subset.id/index.php/IJCSR

FN (False Negative)	TN (True Negative)

Akurasi

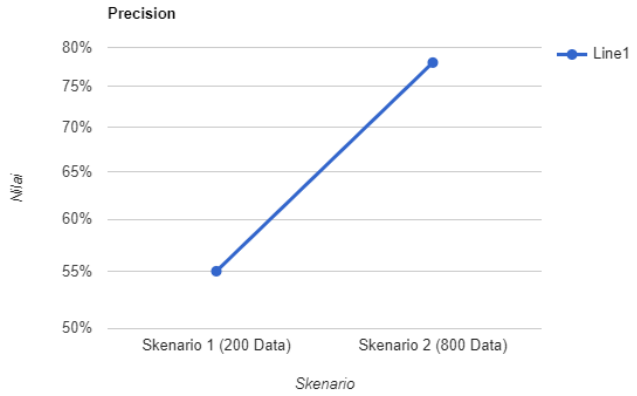
Tingkat akurasi dari hasil analisis dengan algoritma Decision Tree dari kedua scenario menunjukan akurasi meningkat dimana pada scenario 1 (200 data) menunjukan akurasi 58% dan scenario 2 (800 data) dengan akurasi tertinggi yakni 78%.



GAMBAR 17. Akurasi Decision Tree

Precision

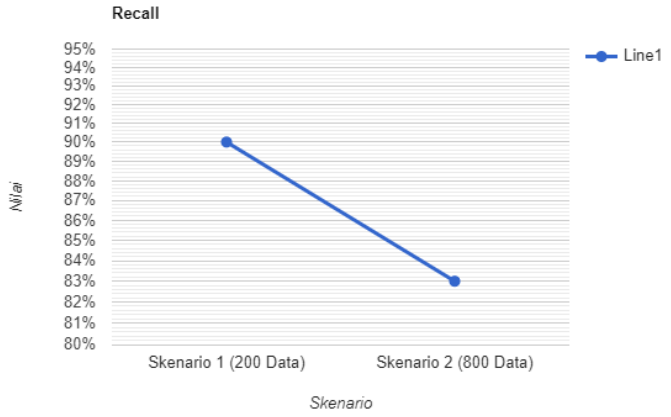
Hasil analisis menunjukan nilai precisioin pada algoritma Decision Tree dimana nilai tersebut menunjukan keakuratan klasifikasi dengan 0.55 pada Skenario 1 dan nilai tertinggi 0.78 pada Skenario 2.



Gambar 18 Precision Decision Tree

Recall

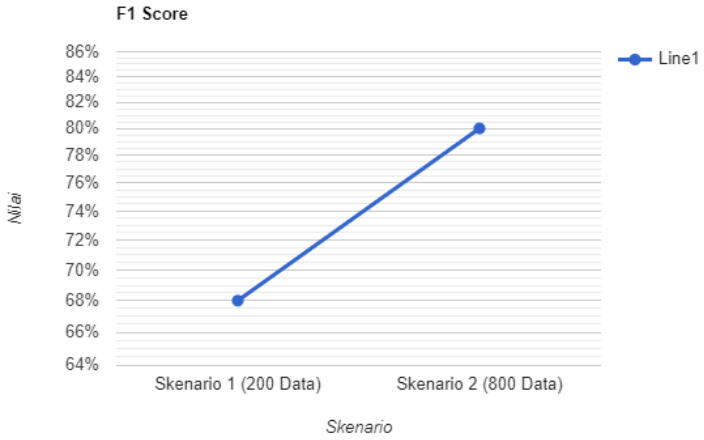
Hasil analisis untuk nilai recall menunjukan apabila semakin tinggi nilai tersebut maka tingkat keberhasilan analisis untuk mengenali kembali suatu dokumen semakin baik. Nilai recall untuk Skenario 1 adalah 0.9 dan untuk scenario 2 memiliki nilai recall rendah yaitu 0.83



GAMBAR 19. Recall Decision Tree

F1 Score

Hasil analisis Skenario 1 0.68 dan Skenario 2 0.8



GAMBAR 20. F1 Score Decision Tree

Perbandingan Naïve Bayes dan Decision Tree

Table 11. hasil perbandingan kedua algoritma

Skenario	Data	Akurasi	Precision	Recall	F1 Score
NB 1	200	78%	0.71	0.84	0.77
NB 2	800	85%	0.83	0.93	0.88
DT 1	200	58%	0.55	0.90	0.68
DT 2	800	78%	0.78	0.83	0.80

Akurasi

Kedua algoritma memiliki nilai akurasi tertinggi pada jumlah data terbanyak dengan 85% pada Naïve Bayes dan 78% pada Decision Tree. Naïve Bayes mengalami kenaikan akurasi sebesar 9% Ketika dataset diganti dengan dataset yang lebih besar sedangkan Decision Tree mengalami lonjakan yang drastis sebesar 34% ketika menggunakan dataset yang lebih besar.

Precision

Nilai Precision pada algoritma Naïve Bayes pada kedua scenario menunjukan angka yang cukup besar yang artinya algoritma dapat memberikan jawaban dengan cukup baik dimana jawaban tersebut adalah bagaimana algoritma tersebut mengklasifikasi sebuah data. Decision Tree memiliki nilai presisi yang kurang baik selebihnya pada skenaria 1 dengan 200 data dimana nilai presisi cukup rendah.

Recall

Seluruh scenario dari kedua algoritma menunjukan nilai yang baik dengan 0.93 pada Naïve Bayes scenario 2 dan 0.90 pada Decision Tree scenario 1. Nilai tersebut menjadi alat ukur untuk mengetahui keberhasilan algoritma dalam menemukan Kembali informasi

F1-Score

Kedua algoritma menunjukan skor yang baik untuk skenario dengan jumlah data lebih besar dengan Naïve Bayes sebesar 0.88 dan Decision Tree sebesar 0.80

KESIMPULAN

Berdasarkan dari hasil pengujian pada bab sebelumnya dapat dibuat kesimpulan bahwa penggunaan dan implementasi algoritma Naïve Bayes dan Decision Tree pada analisis sentimen dengan data kecil memiliki akurasi yang cenderung rendah terutama pada Decision Tree. Perbandingan kedua algortima mendapatkan nilai akurasi tertinggi sebesar 85% untuk Naïve Bayes dan 78% untuk Decision Tree

TINJAUAN PUSTAKA

[1] S. A. Assaidi and F. Amin, "Analisis Sentimen Evaluasi Pembelajaran Tatap Muka 100 Persen pada Pengguna Twitter menggunakan Metode Logistic Regression," *J. Pendidik. Tambusai*, vol. 6, no. 2, 2022, doi: 10.31004/jptam.v6i2.4543.

[2] A. Rozaq, Y. Yunitasari, K. Sussolaikah, E. Resty, N. Sari, and R. I. Syahputra, "Analisis Sentimen Terhadap Implementasi Program Merdeka Belajar Kampus Merdeka Menggunakan Naïve Bayes , K-Nearest Neighboars Dan Decision Tree," *J. Media Inform. Budidarma*, vol. 6, no. April, pp. 746–750, 2022, doi: 10.30865/mib.v6i2.3554.

[3] F. D. Adinata and J. Arifin, "Klasifikasi Jenis Kelamin Wajah Bermasker Menggunakan Algoritma Supervised Learning," *J. Media Inform. Budidarma*, vol. 6, no. 1, p. 229, 2022, doi: 10.30865/mib.v6i1.3377.

[4] S. Samsir, A. Ambiyar, U. Verawardina, Fi. Edi, and R. Watrianthos, "Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naïve Bayes," *J. Media Inform. Budidarma*, vol. 5, no. 1, p. 149, 2021, doi: 10.30865/mib.v5i1.2604.

[5] W. Yulita *et al.*, "Analisis Sentimen Terhadap Opini Masyarakat Tentang Vaksin Covid-19 Menggunakan Algoritma Naïve Bayes Classifier," *Jdmsi*, vol. 2, no. 2, pp. 1–9, 2021.

[6] M. I. Fikri, T. S. Sabrila, and Y. Azhar, "Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter," *Smatika J.*, vol. 10, no. 02, pp. 71–76, 2020, doi: 10.32664/smatika.v10i02.455.

[7] R. Puspita and A. Widodo, "Perbandingan Metode KNN, Decision Tree, dan Naïve Bayes Terhadap Analisis

Sentimen Pengguna Layanan BPJS," *J. Inform. Univ. Pamulang*, vol. 5, no. 4, p. 646, 2021, doi: 10.32493/informatika.v5i4.7622.

[8] K. A. Rokhman, B. Berlilana, and P. Arsi, "Perbandingan Metode Support Vector Machine Dan Decision Tree Untuk Analisis Sentimen Review Komentar Pada Aplikasi Transportasi Online," *J. Inf. Syst. Manag.*, vol. 3, no. 1, pp. 1–7, 2021, doi: 10.24076/joism.2021v3i1.341.

[9] D. Darwis, N. Siskawati, and Z. Abidin, "Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Review Data Twitter BMKG Nasional," *J. Tekno Kompak*, vol. 15, no. 1, p. 131, 2021, doi: 10.33365/jtk.v15i1.744.

[10] N. R. Robynson and Y. Sibaroni, "Analisis Tren Sentimen Masyarakat Terhadap Pembatasan Sosial Berskala Besar Kota Jakarta Menggunakan Algoritma Support Vector Machine," *e-Proceeding Eng.*, vol. 8, no. 5, pp. 10166–10178, 2021.

[11] H. Sujadi, S. Fajar, and C. Roni, "Analisis Sentimen Pengguna Media Sosial Twitter Terhadap Wabah Covid-19 Dengan Metode Naive Bayes Classifier Dan Support Vector Machine," *INFOTECH J.*, vol. 8, no. 1, pp. 22–27, 2022, doi: 10.31949/infotech.v8i1.1883.