

Sentiment Analysis Of Student Opinion Related To Online Learning Using Naïve Bayes Classifier Algorithm And SVM With Adaboost On Twitter Social Media

Analisis Sentimen Terhadap Opini Mahasiswa Terkait Pembelajaran Daring Dengan Menggunakan Algoritma *Naïve Bayes Classifier* Dan SVM Dengan Adaboost Pada Media Sosial Twitter

Mohammad Rizal Ramli¹, Heni Sulastr², Rianto³

Informatika, Universitas Siliwangi Tasikmalaya, Indonesia

¹187006028@student.unsil.ac.id, ^{2*}henisulastr²@unsil.ac.id, ³rianto@unsil.ac.id

*: *Penulis korespondensi (corresponding author)*

Informasi Artikel

Received: Januari 2023

Revised: May 2023

Accepted: May 2023

Published: June 2022

Abstract

Purpose: Twitter is one of the social media that functions to express opinions on issues or problems that are currently happening, such as problems in the social, economic, educational and other fields. One of the issues being discussed so far is online learning. The government has issued a policy, one of which is for all students to study at home online by using a network to be able to interact with each other like in the classroom. The government's reason for issuing this policy is to break the chain of the spread of the Covid-19 virus, which until now has not subsided. Regarding this online learning policy, there are pros and cons. This opinion is widely expressed on social media, one of which is Twitter. Sentiment analysis is a method for analyzing an opinion which aims to classify texts. The Naïve Bayes Classifier and Support Vector Machine methods are methods machine learning that can be used for sentiment analysis. The problem in classifying text is that the resulting accuracy is less than optimal, so feature selection or boosting is needed to improve its accuracy. In this study, optimization of boosting was carried out using Adaboost. The purpose of this study is to compare the performance of the algorithm before and after using Adaboost. The results of the sentiment analysis on online learning obtained the highest accuracy results by the Naïve Bayes Classifier algorithm coupled with Adaboost of 99.26%, with a precision of 99.39% and recall of 99.20%.

Abstrak

Keywords: adaboost; sentiment analysis; covid-19; online lecture; Naïve Bayes; svm; twitter

Kata kunci: adaboost; analisis sentimen; covid-19; kuliah online; Naïve Bayes; svm; twitter

Tujuan: Twitter merupakan salah satu media sosial yang berfungsi untuk mengungkapkan pendapat mengenai isu atau masalah yang sedang terjadi seperti halnya masalah dibidang sosial, ekonomi, pendidikan dan masalah lainnya. Salah satu masalah yang sedang diperbincangkan sampai saat ini yaitu pembelajaran daring. Pemerintah mengeluarkan kebijakan salah satunya kepada seluruh mahasiswa untuk belajar di rumah saja secara *online* dengan menggunakan jaringan untuk bisa saling berinteraksi layaknya di dalam kelas. Alasan pemerintah mengeluarkan kebijakan tersebut untuk memutus rantai penyebaran virus *Covid-19* yang sampai saat ini masih belum reda. Mengenai kebijakan pembelajaran daring ini menimbulkan pro dan kontra. Opini tersebut banyak diungkapkan di media sosial salah satunya media sosial twitter. Analisis sentimen merupakan sebuah metode untuk menganalisis sebuah pendapat yang bertujuan untuk dilakukan mengklasifikasikan teks. Metode *Naïve Bayes Classifier* dan *Support Vector Machine* merupakan metode *machine learning* yang bisa digunakan untuk analisis sentimen. Masalah dalam pengklasifikasian teks adalah kurang optimalnya akurasi yang dihasilkan, sehingga diperlukan seleksi fitur atau *boosting* untuk bisa meningkatkan akurasi. Pada penelitian ini dilakukan optimalisasi *boosting* dengan menggunakan Adaboost. Tujuan dari penelitian ini adalah membandingkan performa algoritma sebelum dan sesudah menggunakan Adaboost. Hasil analisis sentimen terhadap pembelajaran daring ini didapatkan hasil akurasi tertinggi oleh algoritma *Naïve Bayes Classifier* ditambah dengan Adaboost sebesar 99.26%, dengan presisi sebesar 99.39% dan *recall* 99.20%.

1. Pendahuluan

Penyakit *Covid-19* yang berasal dari Wuhan, Tiongkok pada tahun 2020 sampai dengan sekarang sedang marak menyerang tubuh manusia diberbagai belahan dunia [1]. Salah satunya negara Indonesia. Penyakit yang sangat cepat penyebarannya ini umumnya menyerang bagian sistem pernapasan manusia yang pahitnya menyebabkan kematian [2].

Penyakit *Covid-19* ini sangat berdampak terhadap ekonomi, sosial, pendidikan dan lain-lain [3]. Salah satu hal yang sangat penting dan diperhatikan pada saat pandemi seperti ini adalah pendidikan. Pemerintah mengeluarkan kebijakan, yaitu menyuruh kepada masyarakat untuk melakukan Pembatasan Sosial Berskala Besar (PSBB) yang membuat bidang pendidikan harus melaksanakan pembelajaran secara Dalam Jaringan atau Daring [4]. Pembelajaran daring ini

dilakukan sampai sekarang yang akhirnya *Covid-19* ini berdampak kepada pelajar maupun mahasiswa yang diharuskan untuk beraktivitas di rumah masing-masing agar bisa memutus rantai penyebaran *Covid-19* [5].

Dengan dilakukannya pembelajaran secara daring, jadi perbincangan mengenai pembelajaran secara daring ini yang banyaknya dikeluhkan oleh mahasiswa diberbagai media sosial salah satunya ialah Twitter. Dengan aplikasi Twitter ini pengguna akan lebih mudah untuk mengikuti tren, informasi terbaru atau yang masih hangat diperbincangkan dan masih banyak hal lainnya yang bermanfaat pada aplikasi Twitter ini. Pendapat dari seseorang tersebut bisa dimanfaatkan untuk dilakukan analisis sentimen terhadap kasus yang sedang viral. Salah satu topik yang sedang viral dan ramai diperbincangkan, yaitu kebijakan dari pemerintah berupa pembelajaran secara daring [6].

Analisis sentimen ini dikenal juga sebagai *opinion mining* yang merupakan studi komputasi dari pendapat, sikap dan emosi yang disajikan dalam sebuah teks [7]. Analisis sentimen ini bisa digunakan sebagai salah satu cara untuk mengevaluasi terhadap layanan yang telah diberikan dan dilakukan dengan berbagai cara, yaitu diantaranya mengumpulkan pendapat banyak orang di media sosial [8]. Salah satu teknik atau metode yang terdapat dalam text mining untuk analisis sentimen, yaitu *Naïve Bayes Classifier* (NBC) dan *Support Vector Machine* (SVM).

Metode *Naïve Bayes Classifier* sendiri merupakan salah satu *machine learning* yang juga merupakan algoritma untuk mengklasifikasikan sebuah data. Metode *Naïve Bayes* ini metode yang sangat cocok untuk model *Classifier probabilistic* [9]. Metode ini sangat berpotensi baik dalam klasifikasi, presisi dan komputasi data. Metode ini juga banyak digunakan untuk teknik klasifikasi terutama pada media sosial Twitter menggunakan metode seperti unigram *Naïve Bayes*, *Multinomial Naïve Bayes*, dan *maximum entropy classification* [10]. Perhitungan kategori probabilitas pada metode ini menggunakan pendekatan algoritma *bayes* dengan menggunakan beberapa persamaan [11]. Penelitian yang dilakukan oleh Veny, dkk dalam penelitiannya menggunakan algoritma *Naïve Bayes Classifier*, Decision Tree dan Random Forest ini menunjukkan akurasi sebesar 86,43% dari pengujian data menggunakan Naive Bayes yang akurasinya lebih tinggi dibandingkan dengan algoritma lainnya yaitu Decision Tree dan Random Forest yaitu 82,91% [12].

Algoritma SVM umumnya lebih banyak diimplementasikan untuk melakukan klasifikasi yang nantinya akan mencari *hyperplane* yang terbaik dengan memaksimalkan jarak antar kelas yang di mana *hyperplane* ini berfungsi untuk pemisah antar kelas pada suatu proses klasifikasi [13]. Penelitian yang dilakukan oleh Bhavitha, dkk yang dalam penelitiannya menghasilkan akurasi 85%. Support Vector Machine memberikan akurasi yang sangat baik dibandingkan dengan banyak algoritma lainnya [14].

Penelitian yang dilakukan oleh Ankita, dkk yang dalam penelitiannya menghasilkan akurasi yang baik. Teknik klasifikasi yang digunakan adalah pendekatan ensemble seperti AdaBoost yang menggabungkan beberapa *Classifier* lain untuk membentuk satu *Classifier* yang kuat dan memberikan akurasi sebesar 84,5% [15].

Algoritma dari *machine learning* yaitu salah satunya *Naïve Bayes Classifier* dan algoritma *Support Vector Machine* akan dilakukan perbandingan manakah algoritma yang optimal dan memiliki akurasi yang baik dengan ditambahkan metode *boosting* yaitu Adaboost agar bisa lebih meningkatkan akurasi dari kedua algoritma tersebut [16].

Latar belakang tersebut menjadi dasar diperlukannya sebuah analisis sentimen untuk melihat bagaimana respon dan opini mahasiswa indonesia terhadap aktivitas secara daring selama *Covid-19* dengan menggunakan data yang bersumber dari media sosial Twitter yang nantinya akan dilakukan klasifikasi menggunakan algoritma *machine learning Naïve Bayes Classifier* dan algoritma *Support Vector Machine* dengan menggunakan fitur *boosting* yaitu Adaboost.

2. Metode Penelitian

Metode penelitian merupakan sebuah langkah-langkah serta konsep agar bisa memperoleh data yang telah diproses menjadi sebuah informasi yang lebih jelas, akurat dan terperinci. Mulai dari tahapan pengumpulan data, filter data, pelabelan, analisis algoritma sampai analisis hasil algoritma. Pada Gambar 1 terdapat beberapa langkah penelitian seperti tahap awal, *filter data*, label dan *terms*, analisis algoritma dan analisis hasil.

2.1. Tahap Awal

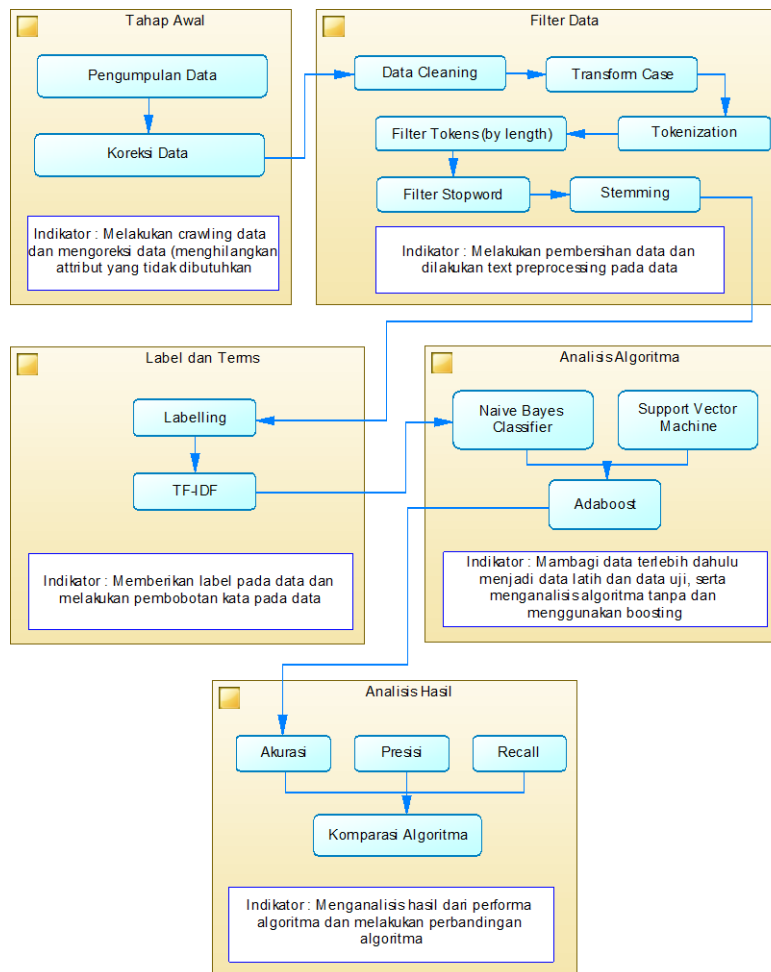
Pada tahapan awal ini dilakukan pengumpulan data terlebih dahulu agar mendapatkan data set. Pengumpulan data menggunakan sebuah teknik *crawling* pada media sosial twitter berdasarkan kata kunci “kuliah daring” dan “kuliah *online*” menggunakan fitur Twitter API. Twitter API ini dapat memudahkan dalam memperoleh data seperti halnya data pengguna, *tweet*, tanggal *tweet* dan lain-lain [17]. Koreksi data merupakan pembuangan atribut yang tidak penting dan atribut yang sama agar data bisa langsung diolah pada tahap *preprocessing*.

2.2. Filter Data

Text preprocessing berguna untuk menyeragamkan sebuah teks serta formatnya agar data bisa dipersiapkan dan diolah pada tahap yang selanjutnya [18]. Terdapat beberapa tahapan pada *text preprocessing* seperti halnya *data cleaning* yang akan membersihkan data seperti halnya menghilangkan tanda baca dan lain-lain. *Transform case* untuk memeriksa ukuran setiap karakter dari awal sampai akhir dan merubah semua karakter pada suatu kalimat yang tadinya kapital menjadi huruf kecil (*lowercase*). *Tokenization* yang nantinya dilakukan pemecahan string atau pun input terhadap suatu teks dan menghapus kata-kata yang tidak penting seperti halnya URL, mention, hastag serta delimiter dan lain-lain. *Filter tokens (by length)* berguna untuk menghilangkan kata yang mengganggu dan bersifat singkatan berdasarkan minimal panjang karakter pada setiap kata dengan minimal 4 karakter pada setiap katanya. *Filter stopwords* untuk menghapus kata yang tidak memiliki arti dan kurang penting serta kata yang tidak memiliki makna tersendiri berdasarkan data *stopword*. *Stemming* mencari kata dasar untuk merubah kata-kata menjadi kata yang berbentuk dasar dan menghilangkan semua imbuhan pada setiap kata.

2.3. Label dan Terms

Pada tahapan ini, yaitu pengambilan label dari data yang telah dipersiapkan sebelumnya menggunakan *text polarity* yang nantinya akan menghasilkan kelas sentimen berupa positif, netral atau pun negatif dengan menggunakan library dari *python* yaitu *Textblob*. Selanjutnya melakukan pembobotan dengan mengubah data tekstual menjadi data numerik menggunakan TF-IDF (*Term Frequency – Inverse Document Frequency*) yang bertujuan untuk menghitung bobot pada suatu kata yang akan digunakan sebagai fitur.



Gambar 1. Metode Penelitian

2.4. Analisis Algoritma

Pada tahapan ini akan dilakukan analisis algoritma untuk mengolah datanya. Pengolahan datanya menggunakan algoritma *Naive Bayes Classifier* dan *Support Vector Machine* dengan Adaboost. Setelah pembobotan kata, dilakukan pembagian data menjadi data latih (*training*) 80% dan data uji (*testing*) 20% dilakukan pada 4 proses yaitu diantaranya *Naive Bayes Classifier* tanpa Adaboost, *Support Vector Machine* tanpa Adaboost, *Naive Bayes Classifier* dengan Adaboost dan *Support Vector Machine* dengan Adaboost.

2.5. Analisis Hasil

Pada tahapan ini akan menampilkan hasil dari performa algoritma yang digunakan, yaitu berupa akurasi yang didapatkan, presisi, *recall* dan akan dilakukan juga evaluasi model serta prediksi data yang telah diklasifikasi menggunakan algoritma *Naive Bayes Classifier* dan *Support Vector Machine* dengan Adaboost yang nantinya akan dilakukan perbandingan atau komparasi algoritma.

3. Hasil dan Pembahasan

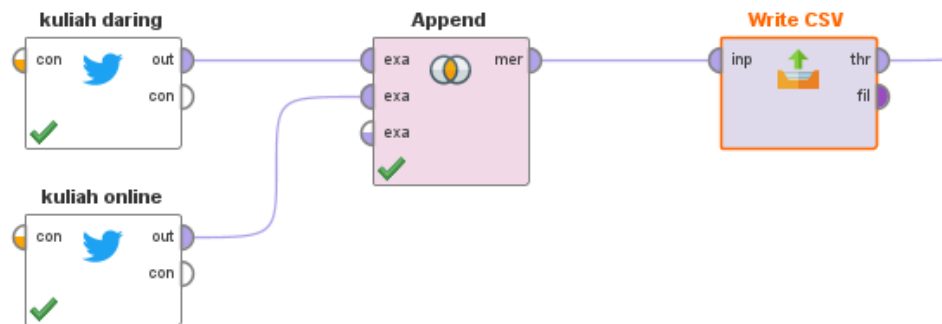
Pada tahap awal penelitian dilakukan pengumpulan data terlebih dahulu menggunakan teknik *crawling data* dengan menggunakan API pada media sosial Twitter.

3.1. Tahap Awal

Pada tahap awal ini terdiri dari tahapan 2 tahapan yaitu tahapan pengumpulan data dan koreksi data.

3.1.1. Pengumpulan Data

Pada tahapan pengumpulan data ini menggunakan tools di Rapidminer seperti pada Gambar 2 berikut.



Gambar 2. Pengumpulan Data

Pada Gambar 2 proses pengumpulan data menggunakan Twitter API dengan operator Search Twitter. Data *tweet* yang diambil yang mengandung kata “kuliah daring” dan “kuliah *online*”.

3.1.2. Data Koreksi

Data yang telah diperoleh pada proses *crawling* akan dikonversikan ke dalam bentuk file CSV. Data pada bulan Mei 2022 dan bulan Juni 2022 akan digabungkan menjadi satu file yang menghasilkan data sebanyak 2909 *tweet* pada bulan Mei 2022 dan 2145 *tweet* pada bulan Juni 2022 dengan *tweet* berbahasa Indonesia seperti pada Gambar 3 berikut.

...	Id	Text
1	1542108334...	@Unsoedfess1963 Setuju banget, tahun sebelumnya bisa ngajuin beberapakali
2	1542063129...	[TATA CARA UNTUK MENGAKSES PERPUSTAKAAN ONLINE DI FIKES UPNVJ]
3	1541990743...	@LndMuslitaNazir pas SMA dikira anak kuliahan, skrg udh jd anak kuliahan dikira br masuk SMA.
4	1541961301...	@scorpiopenyakit OSPEK OFFLINE BERARTI GUE KULIAH LURIBG DOBG...? (sudah terlalu nyam...
5	1541957814...	#INFO: Bimbingan Pra Krs Ibu Anastasia Susty A, dilaksanakan pada hari Kamis, 30 Juni 2022 pkl. ...
6	1541924272...	Rejekinya Shea pas pandemi :
7	1541805002...	#dnss2022
8	1541767237...	Ig menikmati masa masa liburan, kl udh masuk kuliah kepala cepet panas soalnya, daring aja pana...
9	1541759256...	@efener dari sebelum pandemi, Binus sudah develop Kuliah Daring, begitu online gini, lancar jaya ...
10	1541749868...	@collegemenfess Wajar aja sih, kuliah daring udh gtu rumahnya jauh dri kampus. Apalagi yg kuran...

Gambar 3. Hasil Tahap Koreksi Data

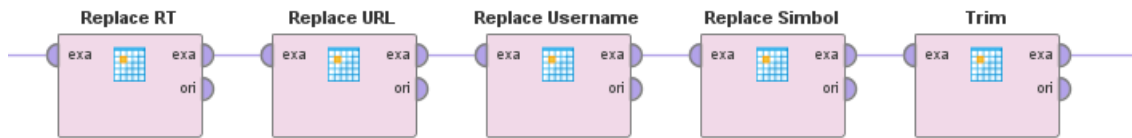
Pada Gambar 3 menghapus beberapa atribut yang tidak penting seperti halnya tanggal *tweet*, nama pengguna, bahasa serta menghilangkan data yang bersifat duplikat dan hasil akhir data menjadi 2685 data.

3.2. Filter Data

Pada tahap filter data ini terdiri dari beberapa tahapan seperti halnya *cleaning data*, *Transform Case*, *Tokenization*, filter tokens (by length), filter *Stopwords* dan *Stemming*.

3.2.1. Data Cleaning

Proses *Data Cleaning* ini terdiri dari beberapa tahap atau operator yang di mana tiap operator itu berbeda-beda fungsinya.



Gambar 4. Filter Data

Pada Gambar 4 proses ini dilakukan pembersihan data dari *retweet* pengguna, *link* URL atau *link* web, *username*, simbol yang tidak penting dan menghapus spasi yang berlebihan. Hasil dari tahapan di atas bisa dilihat pada **Tabel 1** berikut.

Tabel 1. Proses *Data Cleaning*

Sebelum Data Cleaning	Sesudah Data Cleaning
RT @hopbii: @convomfs Kuliah daring planga plongo pas skripsi langsung terjungkang	Kuliah daring planga plongo pas skripsi langsung terjungkang

Pada Tabel 1 proses *Data Cleaning* ini ditampilkan sebelum dan sesudah prosesnya. Huruf yang diberi warna merah akan dihilangkan seperti simbol, *link*, *username* dan *Retweet*.

3.2.2. Transform Case

Transform Case merupakan sebuah proses untuk mengkonversi sebuah teks yang bersifat kapital atau huruf besar (*uppercase*) menjadi huruf kecil dan sama rata (seragam) atau bisa disebut *lowercase*. Hasil tahap *Transform Case* bisa dilihat pada Tabel 2 berikut.

Tabel 2. Proses *Transform Case*

Sebelum Transform Case	Sesudah Transform Case
OSPEK OFFLINE BERARTI GUE KULIAH LURIBG DOBG sudah terlalu nyaman dengan daring	ospek offline berarti gue kuliah luribg dobg sudah terlalu nyaman dengan daring

Pada Tabel 2 mengubah huruf kapital menjadi huruf kecil salah satunya huruf awal yang kebanyakan huruf besar.

3.2.3. Tokenization

Pada proses ini *Tokenization* ini merupakan pemisahan/pemotongan tiap kata yang terdapat pada sebuah dokumen menjadi potongan tunggal. Berikut merupakan hasil *Tokenization*.

Tabel 3. Proses *Tokenization*

Sebelum Tokenization	Sesudah Tokenization
aku yg harusnya mau kerja jadi off selama pandemi jadi fulltime mom	'Aku', 'yg', 'harusnya', 'mau', 'kerja', 'jadi', 'off', 'selama', 'pandemi', 'Jadi', 'fulltime', 'mom'

Hasil dari proses pada Tabel 3 tersebut adalah berupa beberapa kalimat yang dipisahkan menjadi per kata

3.2.4. Filter Tokens (by Lengths)

Tahapan ini merupakan proses penghapusan kata berdasarkan panjang karakter yang telah ditentukan yaitu minimal 4 huruf dan maksimal 24 dari panjangnya kata pada dokumen. Hal ini dikarenakan untuk mengurangi *typo*. Hasil ini bisa dilihat pada Tabel 4 berikut.

Tabel 4. Proses *Filter Tokens*

Sebelum Filter	Sesudah Filter
'Aku', 'yg', 'harusnya', 'mau', 'kerja', 'jadi', 'off', 'selama', 'pandemi', 'Jadi', 'fulltime', 'mom'	'harusnya', 'kerja', 'jadi', 'selama', 'pandemi', 'Jadi', 'fulltime'

Hasil dari proses *filter tokens* pada Tabel 4 diatas dengan huruf berwarna merah pada tabel sebelum filter akan dihilangkan, karena mengandung huruf yang kurang dari kriteria yaitu 4 huruf.

3.2.5. Filter Stopwords

Tahap ini merupakan sebuah proses penghapusan kata yang tidak memiliki makna atau arti berdasarkan data dari dokumen *Stopwords*. Pada proses ini menggunakan dokumen yang telah disiapkan sebelumnya pada file 'stopwordbahasa.csv' sebagai kamus atau patokan yang berisi kata-kata umum dalam jumlah yang banyak.

Tabel 5. Proses *Filter Stopwords*

Sebelum Stopwords	Sesudah Stopwords
'semenjak', 'kuliah', 'daring', 'kemeja', 'jualan', 'semua', 'taunya', 'bahkan', 'jadi', 'trend'	'semenjak', 'kuliah', 'daring', 'kemeja', 'jualan', 'taunya', 'trend'

Pada Tabel 5 hasil dari proses *Stopwords* tersebut adalah menghapus kata-kata yang tidak memiliki arti atau makna dan ditandai dengan warna merah pada bagian sebelum *Stopwords*

3.2.6. Stemming

Tahap ini merupakan proses perubahan kata menjadi kata ke bentuk kata dasar. Hasil dari proses *Stemming* bisa dilihat pada Tabel 6 berikut.

Tabel 6. Proses *Stemming*

Sebelum Stemming	Sesudah Stemming
'Kuliah', 'full', 'daring', 'kerjaan', 'santai', 'Yuhuu'	['kuliah'], ['full'], ['daring'], ['kerja'], ['santai'], ['yuhuu']

Seperti pada Tabel 6 di atas hasil dari proses tersebut adalah mengubah kata yang berimbuhan menjadi kata dasarnya.

3.3. Label dan Terms

Pada tahap pelabelan akan diberi kelas pada tiap sentimen menjadi kelas positif, negatif dan netral. Setelah itu akan dihitung bobotnya dengan menggunakan TF-IDF yang berfungsi untuk mengubah data teks menjadi numerik.

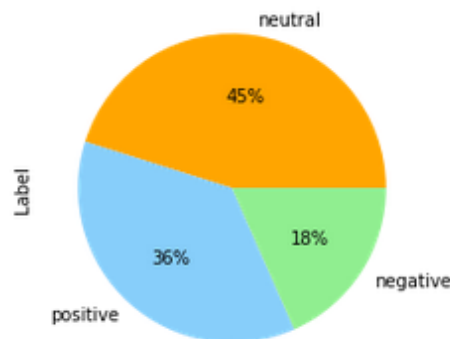
3.3.1. Labelling

Pada proses label ini menggunakan *Library* pada *Python* yaitu *Textblob*. *Textblob* menghitung rata-rata untuk memberikan nilai polaritas dan subjektivitas. Nilai polaritas dengan bilangan tertentu akan diberi label *positive* (Positif) jika polaritas lebih dari 0, *neutral* (Netral) jika polaritas sama dengan 0 dan *negative* (Negatif) jika polaritas kurang dari 0. Berikut merupakan hasil dari tahapan *Labelling*.

	Processed	Polarity	Subjectivity	Label
0	breathe since forget date	0.0	0.0	neutral
1	moment students	0.0	0.0	neutral
2	practitioner series program activity form guest divide several sessions base course offer even semester	-0.4	0.5	negative
3	can ui ui partner pt accept	0.0	0.0	neutral
4	difference receive pt partner pt	0.0	0.0	neutral

Gambar 5. Labelling

Pada Gambar 5 ditampilkan hasil dari polaritas dan label. Data yang telah diberi label oleh *Textblob* sebanyak 2.685 data *tweet* dengan 980 data positif, 1.212 data netral dan 493 data negatif.



Gambar 6. Visualisasi Hasil Pelabelan

Pada Gambar 6 merupakan hasil visualisasi hasil pelabelan dalam bentuk diagram pie dengan jumlah persentase. Label yang paling mendominasi adalah label netral (warna *orange*) dengan jumlah persentase 45%. Sedangkan label positif (warna biru) berjumlah 36% dan label negatif (warna hijau) berjumlah 18%.

3.3.2. TF-IDF

TF-IDF merupakan seleksi fitur yang berfungsi untuk menghitung skor pada frekuensi kemunculan pada setiap kata yang terdapat pada teks atau *tweet*.

Tabel 7. Proses TF-IDF

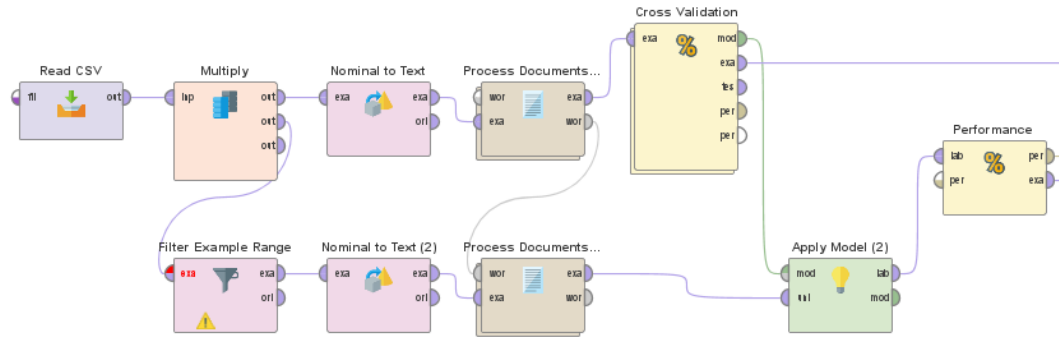
Term	TF			DF	IDF	TF-IDF		
	d1	d2	d3			d1	d2	d3
adek	1	0	0	1	0.477	0.08	0	0
bentar	1	0	0	1	0.477	0.08	0	0
masuk	1	0	0	1	0.477	0.08	0	0
lucu	1	0	0	1	0.477	0.08	0	0
daring	1	1	1	3	0	0	0	0
menerus	1	0	0	1	0.477	0.08	0	0
coret	0	2	0	1	0.477	0	0.136	0
baju	0	1	0	1	0.477	0	0.068	0
bagus	0	1	0	1	0.477	0	0.068	0
pake	0	1	0	1	0.477	0	0.068	0
kuliah	0	1	1	2	0.176	0	0.025	0.029
suruh	0	0	1	1	0.477	0	0	0.08
mudik	0	0	1	1	0.477	0	0	0.08
efektif	0	0	1	1	0.477	0	0	0.08
semangat	0	0	1	1	0.477	0	0	0.08
Panjang Dokumen	6	7	6					

Keterangan:

- TF = Jumlah kemunculan kata atau term pada sebuah dokumen
- d1 = Dokumen 1 dengan kalimat “adek bentar masuk lucu daring menerus”
- d2 = Dokumen 2 dengan kalimat “coret baju bagus pake kuliah daring”
- d3 = Dokumen 3 dengan kalimat “suruh mudik efektif semangat kuliah daring”
- DF = Jumlah kemunculan kata atau term pada seluruh dokumen
- IDF = Hasil pembagian jumlah dokumen dan DF dengan logaritma
- TF-IDF = Hasil perkalian dari TF dan IDF

3.4. Analisis Algoritma

Pada tahap ini menganalisis dengan menggunakan algoritma *Naïve Bayes Classifier* dan Algoritma *Support Vector Machine* (SVM) yang ditambah dengan metode *boosting* yaitu *Adaptive Boosting* (Adaboost). Pada tahap ini dilakukan pembagian data terlebih dahulu menjadi 80% data *training* dan 20% data *testing*.



Gambar 7. Proses Analisis Algoritma

Proses pada **Gambar 7** merupakan proses untuk menganalisis performa pada tiap algoritma yaitu *Naïve Bayes* tanpa dan menggunakan Adaboost, *SVM* tanpa dan menggunakan Adaboost.

3.5. Analisis Hasil

Pada tahapan ini menggunakan *Confusion Matrix* yang merupakan sebuah pengukuran performa pada *machine learning* untuk klasifikasi lebih dari 3 kelas yang dapat dilihat pada Tabel 8 berikut.

Tabel 8. *Confusion Matrix*

Confusion Matrix		True Class		
		True Positive	True Negative	True Neutral
Predicted Class	Pred Positive	TP	FP	FN
	Pred Negative	FN	TN	FN
	Pred Neutral	FNeu	FNeu	TNeu

Keterangan:

- TP (*True Positive*) = Jumlah prediksi yang benar dari data yang relevan.
- TN (*True Negative*) = Jumlah prediksi yang benar dari data yang relevan.
- TNeu (*True Neutral*) = Jumlah prediksi yang benar dari data yang relevan.
- FP (*False Positive*) = Jumlah prediksi yang salah dari data yang relevan.
- FN (*False Negative*) = Jumlah prediksi yang salah dari data yang relevan.
- FNeu (*False Neutral*) = Jumlah prediksi yang salah dari data yang relevan.

Terdapat rumus dari akurasi, presisi dan *recall* yaitu diantaranya:

- Akurasi

$$\frac{TP+TN+TNeu}{TP+TN+TNeu+FP+FN+FNeu} \quad (1)$$

- Presisi

$$\frac{TP}{TP+FP} \quad (2)$$

3. Recall

$$\frac{TP}{TP+FN+FNeu} \quad (3)$$

Menghitung akurasi, presisi dan *recall* pada algoritma *Naïve Bayes Classifier* dan SVM itu mekanismenya sama saja. Hanya hasil akhir lah yang akan membedakan diantara keduanya mana yang lebih baik.

Berikut merupakan 4 hasil perbandingan performa algoritma sebelum dan sesudah menggunakan Adaboost.

1. Naïve Bayes Classifier tanpa Adaboost

Pada algoritma *Naïve Bayes Classifier* tanpa menggunakan Adaboost mendapatkan akurasi sebesar 88.83%, dengan presisi 86.82% dan *recall* 90.9%.

2. SVM tanpa Adaboost

Pada algoritma SVM tanpa menggunakan Adaboost mendapatkan akurasi sebesar 88.27%, dengan presisi 92.44% dan *recall* 86.20%.

3. Naïve Bayes Classifier dengan Adaboost

Pada algoritma *Naïve Bayes Classifier* dengan menggunakan Adaboost mendapatkan akurasi sebesar 99.26%, dengan presisi 99.39% dan *recall* 99.20% yang di mana Adaboost meningkatkan akurasi, presisi dan *recall* dibandingkan dengan algoritma *Naïve Bayes Classifier* yang tidak menggunakan Adaboost.

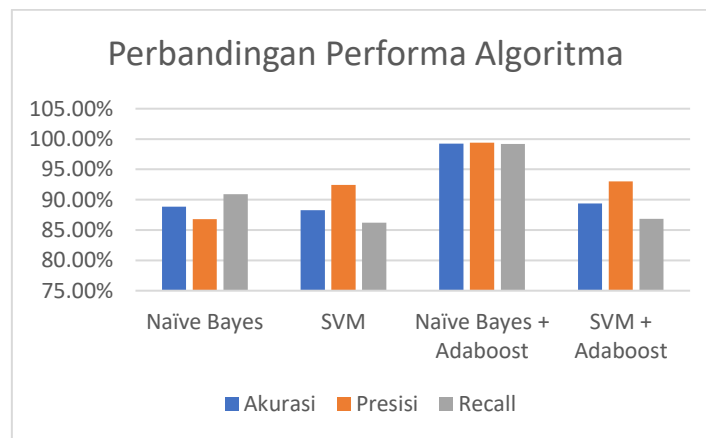
4. SVM dengan Adaboost

Pada algoritma SVM dengan menggunakan Adaboost mendapatkan akurasi sebesar 89.39%, dengan presisi 93.01% dan *recall* 86.87%. Sama halnya dengan *Naïve Bayes*. Penambahan Adaboost ini meningkatkan akurasi, presisi dan *recall* pada algoritma SVM. Hasil perbandingan akurasi, presisi, *recall* dari tiap algoritma dengan menggunakan dan tanpa Adaboost bisa dilihat dari Tabel 9 berikut.

Tabel 9. Hasil Performa Algoritma

Performa Algoritma	Akurasi	Presisi	Recall
<i>Naïve Bayes</i>	88.83%	86.82%	90.90%
SVM	88.27%	92.44%	86.20%
<i>Naïve Bayes</i> + Adaboost	99.26%	99.39%	99.20%
SVM + Adaboost	89.39%	93.01%	86.87%

Pada Tabel 9 dapat disimpulkan bahwa algoritma yang memiliki akurasi tertinggi didapatkan oleh algoritma *Naïve Bayes* + Adaboost sebesar 99.26%, dengan presisi sebesar 99.39% dan *recall* 99.20%. algoritma *Naïve Bayes* merupakan algoritma terbaik dibandingkan dengan algoritma SVM dan metode *boosting* Adaboost terbukti bisa meningkatkan performa bagi tiap algoritma.



Gambar 8. Visualisasi Performa Algoritma

Jika dilihat dari grafik atau visualisasi performa algoritma pada Gambar 8 dapat disimpulkan bahwa algoritma yang memiliki akurasi, presisi dan *recall* tertinggi didapatkan oleh algoritma *Naïve Bayes* + Adaboost dengan grafik tertinggi. Ditambahkannya *boosting* Adaboost sebagai keterbaruan dari penelitian ini yang berfungsi untuk meningkatkan akurasi dari tiap algoritma yang digunakan hasilnya terealisasi. Kedua algoritma yaitu *Naïve Bayes Classifier* dan SVM yang ditambah dengan *boosting* Adaboost meningkat drastis dikarenakan Adaboost membantu meningkatkan kinerja klasifikasi pada tiap algoritma.

4. Kesimpulan dan Saran

Kesimpulan Hasil penelitian dan pengujian dari algoritma *Naïve Bayes Classifier* dan SVM sebelum dan sesudah menggunakan metode *boosting* Adaboost untuk menganalisis opini mahasiswa terkait pembelajaran daring didapatkan akurasi yang paling besar pada algoritma *Naïve Bayes Classifier* ditambah dengan metode *boosting* Adaboost yaitu 99.26%, dengan presisi sebesar 99.39% dan *recall* sebesar 99.20%. Opini mahasiswa terkait pembelajaran daring pada media sosial twitter dengan pengujian algoritma *Naïve Bayes Classifier* dan SVM dengan menggunakan metode *boosting* Adaboost yang paling dominan adalah opini netral. Banyak juga mahasiswa yang beropini positif pada kegiatan pembelajaran daring ini dibanding dengan opini negatif yang artinya sebagian mahasiswa sangat nyaman dalam kegiatan pembelajaran daring. Masih banyak algoritma yang bisa dipadukan dengan metode *boosting* dan masih banyak metode *boosting* yang bisa dipadukan dengan algoritma yang lain untuk bisa mencapai performa yang lebih tinggi.

Daftar Pustaka

- [1] Eben Munthe, "Pandemi Covid-19 dan Dampaknya terhadap Angka Perceraian dan Kelahiran," *Pneum. J. Teol. Kependetaan*, vol. 12, no. 2, pp. 104–115, 2022.
- [2] S. S. Putri Wulandini, "HUBUNGAN PENDIDIKAN DENGAN UPAYA PENCEGAHAN COVID-19 DI DESA LAKSAMANA KECAMATAN SABAK AUH KABUPATEN SIAK," *J. Ilmu Kesehat. Masy. (Journal Public Heal. Sci.)*, vol. 10, 2021.
- [3] I. G. G. P. A. P. Ida Bagus Gede Paramita, "New Normal Bagi Pariwisata Bali Di Masa Pandemi Covid-19," *J. Ilm. Pariwisata Agama dan Budaya*, vol. EISSN 2614, p. ISSN 2527-9734, 2020,

doi: 10.36275/mws.

- [4] A. J. Rajab, M. S. Nurdin, and H. Mubarak, "Tinjauan Hukum Islam pada Edaran Pemerintah dan MUI dalam Menyikapi Wabah Covid-19," *BUSTANUL FUQAHA J. Bid. Huk. Islam*, vol. 1, no. 2, pp. 156–173, 2020.
- [5] Ahmadi, "Pelaksanaan Pembelajaran Di Stai Rakha Sebelum, Semasa Dan Sesudah Pandemi Covid-19," *Adiba J. Educ.*, vol. 2, no. 1, pp. 51–63, 2022.
- [6] A. R. Rizkirobby, M. Nasrun, and R. A. Nugrahaeni, "Deteksi Ujaran Ancaman Berbasis Website Pada Postingan Media Sosial Twitter Menggunakan Metode Naive Bayes," *eProceedings Eng.*, vol. 8, no. 1, pp. 492–499, 2021.
- [7] A. Rahmatulloh, R. N. Shofa, I. Darmawan, and Ardiansah, "Sentiment Analysis of Ojek Online User Satisfaction Based on the *Naive Bayes* and Net Brand Reputation Method," *2021 9th Int. Conf. Inf. Commun. Technol. ICoICT 2021*, pp. 337–341, 2021, doi: 10.1109/ICoICT52021.2021.9527466.
- [8] N. L. P. C. Savitri, R. A. Rahman, R. Venyutzky, and N. A. Rakhmawati, "Analisis Klasifikasi Sentimen Terhadap Sekolah Daring pada Twitter Menggunakan Supervised Machine Learning," *J. Tek. Inform. dan Sist. Inf.*, vol. 7, no. 1, pp. 47–58, 2021, doi: 10.28932/jutisi.v7i1.3216.
- [9] M. I. Ahmadi, D. Gustian, and F. Sembiring, "Analisis Sentiment Masyarakat terhadap Kasus Covid-19 pada Media Sosial Youtube dengan Metode Naive bayes," *J. Sains Komput. Inform. (J-SAKTI)*, vol. 5, no. 2, pp. 807–814, 2021.
- [10] R. Slamet, W. Gata, A. Novtariany, K. Hilyati, and F. A. Jariyah, "Analisis Sentimen Twitter Terhadap Penggunaan Artis Korea Selatan Sebagai Brand Ambassador Produk Kecantikan Lokal," *INTECOMS J. Inf. Technol. Comput. Sci.*, vol. 5, no. 1, pp. 145–153, 2022, doi: 10.31539/intecomsv5i1.3933.
- [11] A. Bayhaqy, S. Sfenrianto, K. Nainggolan, and E. R. Kaburuan, "Sentiment Analysis about E-Commerce from Tweets Using Decision Tree, K-Nearest Neighbor, and *Naive Bayes*," *2018 Int. Conf. Orange Technol. ICOT 2018*, pp. 1–6, 2018, doi: 10.1109/ICOT.2018.8705796.
- [12] V. A. Fitri, R. Andreswari, and M. A. Hasibuan, "Sentiment analysis of social media Twitter with case of Anti-LGBT campaign in Indonesia using *Naive Bayes*, decision tree, and random forest algorithm," *Procedia Comput. Sci.*, vol. 161, pp. 765–772, 2019, doi: 10.1016/j.procs.2019.11.181.
- [13] A. M. Pravina, I. Cholissodin, and P. P. Adikara, "Analisis Sentimen Tentang Opini Maskapai Penerbangan pada Dokumen Twitter Menggunakan Algoritme Support Vector Machine (SVM)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 3, pp. 2789–2797, 2019, [Online]. Available: <http://j-ptiik.ub.ac.id>.
- [14] B. K. Bhavitha, A. P. Rodrigues, and N. N. Chiplunkar, "Comparative study of machine learning techniques in sentimental analysis," *Proc. Int. Conf. Inven. Commun. Comput. Technol. ICICCT 2017*, no. March 2017, pp. 216–221, 2017, doi: 10.1109/ICICCT.2017.7975191.
- [15] A. Rane and A. Kumar, "Sentiment Classification System of Twitter Data for US Airline Service Analysis," *Proc. - Int. Comput. Softw. Appl. Conf.*, vol. 1, pp. 769–773, 2018, doi: 10.1109/COMPSAC.2018.00114.
- [16] N. Novianti, M. Zarlis, and P. Sihombing, "Penerapan Algoritma Adaboost Untuk Peningkatan Kinerja Klasifikasi Data Mining Pada Imbalance Dataset Diabetes," *J. Media Inform. Budidarma*, vol. 6, no. April, pp. 1200–1206, 2022, doi: 10.30865/mib.v6i2.4017.

- [17] A. Syaifuddin *et al.*, “Analisis Sentimen Pada Sosial Media Tentang Implementasi Kebijakan Pse Kominfo Menggunakan Algoritme Lexicon,” *Pros. Semastek 2022*, vol. 1, no. 1, pp. 8–15, 2022.
- [18] F. Bei and S. Sudin, “Analisis Sentimen Aplikasi Tiket Online Di Play Store Menggunakan Metode Support Vector Machine (Svm),” *Sismatik*, vol. 01, no. 01, pp. 91–97, 2021.