

Analisis Sentimen Masyarakat Terhadap Perkuliahan Daring di *Twitter* Menggunakan Algoritma *Naive Bayes* dan *Support Vector Machine*

Federick Dedi Samuel¹, Prima Dina Atika^{1,*}, Siti Setiawati¹

* Korespondensi: e-mail: prima.dina@dsn.ubharajaya.ac.id

¹ Informatika; Fakultas Ilmu Komputer; Universitas Bhayangkara Jakarta Raya; Jl. Raya Perjuangan, Margamulya, Bekasi ;telp.(021) 7231948; e-mail: federick.dedi.samuel19@mhs.ubharajaya.ac.id, prima.dina@dsn.ubharajaya.ac.id, siti.setiawati@dsn.ubharajaya.ac.id

Submitted : **27 September 2023**
Revised : **19 Oktober 2023**
Accepted : **10 November 2023**
Published : **30 November 2023**

Abstract

The COVID-19 pandemic has changed the education landscape around the world, resulting in the cessation of in-person teaching and learning activities and encouraging the adoption of online learning systems. Many Indonesians express their opinions and thoughts about online courses through the social media *Twitter*. Therefore, this study aims to analyze people's sentiment towards online lectures on *Twitter* using *Naive Bayes* and *Support Vector Machine (SVM)* methods. Data for sentiment analysis is taken from *Twitter* using the keywords "#college", "#daring", and "#kuliahdaring". This study limits data collection to the range of 2021-2022. A total of 1,260 Tweets were analyzed, with 633 Tweets having positive sentiments and 627 Tweets having negative sentiments. This study uses *Naive Bayes* and *Support Vector Machine* algorithms to classify positive and negative sentiments in Tweets. The results showed that *Naive Bayes* algorithm achieved the highest accuracy of 72%, while *Support Vector Machine* achieved 66% accuracy.

Keywords: *Naive Bayes*, *Online Course*, *Sentiment Analysis*, *Support Vector Machine*, *Twitter*

Abstrak

Pandemi COVID-19 telah mengubah *landscape* pendidikan di seluruh dunia, mengakibatkan terhentinya kegiatan belajar mengajar secara langsung dan mendorong adopsi sistem pembelajaran daring. Masyarakat Indonesia banyak mengungkapkan opini dan pendapat mereka mengenai perkuliahan daring melalui media sosial *Twitter*. Oleh karena itu, penelitian bertujuan untuk menganalisis sentimen masyarakat terhadap perkuliahan daring di *Twitter* menggunakan metode *Naive Bayes* dan *Support Vector Machine (SVM)*. Data untuk analisis sentimen diambil dari *Twitter* dengan menggunakan kata kunci "#perkuliahan", "#daring", dan "#kuliahdaring". Penelitian ini membatasi pengambilan data pada rentang tahun 2021-2022. Sebanyak 1.260 *Tweet* dianalisis, dengan 633 *Tweet* memiliki sentimen positif dan 627 *Tweet* memiliki sentimen negatif. Penelitian menggunakan algoritma *Naive Bayes* dan *Support Vector Machine* untuk mengklasifikasikan sentimen positif dan negatif dalam *Tweet*. Hasil penelitian menunjukkan bahwa algoritma *Naive Bayes* mencapai akurasi tertinggi sebesar 72%, sementara *Support Vector Machine* mencapai akurasi 66%.

Kata kunci: *Analisis Sentimen*, *Naive Bayes*, *Perkuliahan daring*, *Support Vector Machine*, *Twitter*

1. Pendahuluan

Wabah *COVID-19* telah menimbulkan krisis kesehatan global yang mempengaruhi tingkat Kesehatan di seluruh dunia. Pada 11 Maret 2020, Organisasi Kesehatan Dunia (*WHO*) mengumumkan *COVID-19* sebagai pandemi global. Di Indonesia, untuk mengurangi penyebaran virus melalui kontak fisik, Presiden Joko Widodo memutuskan untuk menerapkan kebijakan Pembatasan Sosial Berskala Besar (*PSBB*) (Dzulfaroh & Naufal, 2021). Penerapan kebijakan *PSBB* di Indonesia telah berdampak luas, termasuk membatasi kegiatan belajar mengajar di semua jenjang pendidikan, dari sekolah dasar hingga perguruan tinggi. Pemerintah Indonesia merespons dengan mengeluarkan himbauan untuk menutup sekolah selama pandemi *COVID-19*. Kementerian Pendidikan dan Kebudayaan mengeluarkan surat edaran yang mewajibkan pembelajaran dilakukan dari rumah melalui metode pembelajaran daring atau jarak jauh, guna memberikan pembelajaran yang lebih bermakna. Pembelajaran daring dilakukan melalui internet dengan menggunakan perangkat seperti *smartphone*, gawai, atau komputer sebagai media pembelajaran. Tujuannya adalah agar siswa dapat belajar tanpa terhambat akses dan dapat mengikuti pembelajaran atau mata kuliah tertentu secara online. Beberapa platform seperti *Zoom*, *Google Classroom*, *Google Meet*, dan lain-lain dimanfaatkan sebagai media pembelajaran (Damanik et al., 2021).

Pembelajaran daring menjadi permasalahan yang harus dihadapi karena harus adanya proses penyesuaian dengan sistem belajar mengajar dalam kegiatan belajar mengajar interaksi mahasiswa dengan dosen sangat penting, baik interaksi mahasiswa dengan dosen maupun interaksi mahasiswa antar mahasiswa harus selalu dibangun untuk membangkitkan semangat belajar sehingga mahasiswa dapat mencapai hasil yang maksimal dan meningkatkan komunikasi serta diskusi tentang setiap kegiatan dalam proses belajar mengajar (Parmono, 2018) karena adanya perubahan yang tiba-tiba dalam sistem pendidikan, proses belajar mengajar yang biasanya dilakukan secara langsung (tatap muka) telah berubah menjadi pembelajaran daring (online). Pembelajaran daring menuai berbagai banyak tanggapan dari masyarakat Indonesia. Jumlah opini yang sangat banyak dan terus-menerus meningkat menjadi permasalahan untuk mengetahui sentimen masyarakat sehingga diperlukan analisis terhadap opini masyarakat (Sidik et al., 2022). Dalam jurnal (Parmono, 2018) disebutkan bahwa mahasiswa yang melakukan pembelajaran daring mengalami kesulitan dalam memahami materi karena pembelajaran yang bersifat satu arah. Dengan pesatnya kemajuan teknologi, banyak orang memilih untuk mengungkapkan pendapat dan ekspresi mereka melalui *media sosial* seperti *youtube*, *tiktok*, *Instagram*, *facebook* serta *Twitter*.

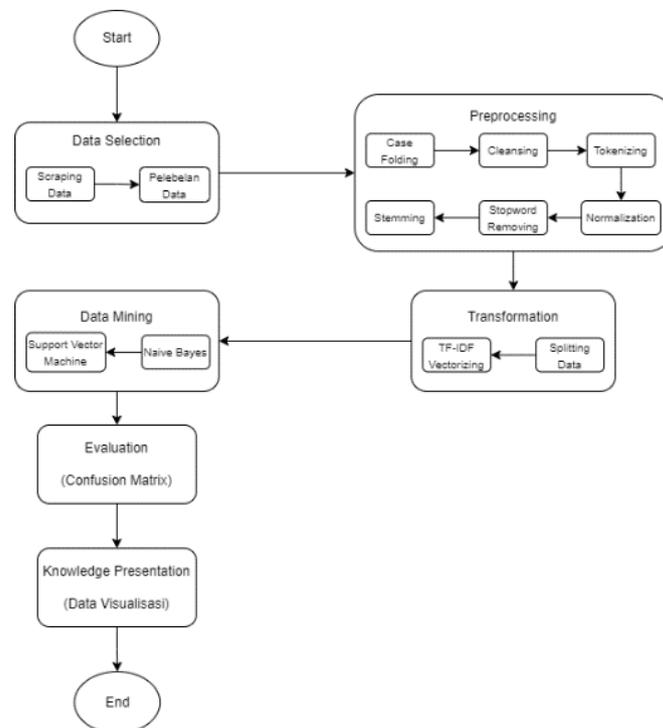
Twitter bisa dijadikan sebagai alat untuk mengevaluasi kepuasan masyarakat terhadap pembelajaran daring, dengan melihat postingan opini dan pendapat mereka dalam tiga kategori sentimen, yaitu positif, netral, dan negatif. Di *platform* tersebut, pengguna dapat memposting opini mereka sehingga dapat dilihat oleh banyak orang. Semua dapat digunakan sebagai media pembelajaran daring agar pembelajaran lebih variatif. Salah satu media sosial yang banyak dipilih orang untuk mengungkapkan pendapat dan ekspresi mereka adalah *Twitter*.

Menurut survei *Twitter* pada tahun 2019, pada tahun sebelumnya jumlah pengguna aktif *Twitter* mencapai 126 juta dengan pertumbuhan pengguna mencapai 9% setiap tahunnya. Hal ini mendorong banyak orang untuk menuangkan opini dan pendapat mereka di *Twitter* agar bisa dilihat oleh banyak orang (Setiawan & Utami, 2021).

Hasil analisis sentimen tersebut bisa digunakan sebagai masukan bagi pemerintah dalam membuat kebijakan di masa depan, khususnya terkait dengan kebijakan pembelajaran daring. Analisis sentimen merupakan salah satu solusi untuk menyaring opini masyarakat dan dapat mengklasifikasikannya ke dalam kelas positif atau kelas negatif. Hasil dari pengklasifikasian data tersebut dapat membantu instansi terkait untuk memberikan tanggapan yang lebih tepat dan strategis kepada setiap masyarakat dan mahasiswa. Analisis sentimen pada penelitian menggunakan data yang diambil dari *Twitter* dengan kata kunci “#perkuliahan”, “#daring”, dan “#kuliahdaring”. Data tersebut akan digunakan sebagai data masukan untuk membangun sebuah model *machine learning* klasifikasi sentimen, model tersebut diharapkan dapat dengan tepat mengklasifikasikan apakah sebuah postingan masyarakat memiliki sentimen negatif dan positif. Peneliti juga membatasi pengambilan data hanya diambil direntang tahun 2021-2023.

2. Metode Penelitian

Penelitian menerapkan metodologi *KDD* untuk melakukan klasifikasi opini masyarakat mengenai Perkuliahan Daring menggunakan metode *Naive Bayes* dan *Support Vector Machine* sehingga dapat menganalisis hasil evaluasi akurasi kedua model. Pemrosesan data dilakukan menggunakan pemrograman *Python* dengan *tools Google Colaboratory*.



Sumber: Hasil Penelitian (2023)

Gambar 1. Alur Penelitian

2.1. Knowledge Discovery in Databases

Knowledge discovery in databases adalah bentuk pembelajaran mesin yang menemukan pengetahuan yang menarik dari basis data dan merepresentasikan pengetahuan tersebut dalam bahasa tingkat tinggi (Ziarko & van Rijsbergen, 1994). *Knowledge discovery in databases* proses mengidentifikasi pola yang valid, baru, berpotensi berguna, dan pola yang pada akhirnya dapat dipahami dalam data (Sánchez, 2013).

2.2. Algoritma Naïve Bayes

Naive Bayes adalah suatu teknik klasifikasi yang menggunakan teorema Bayes untuk menghitung probabilitas kelas dari suatu data (Aggarwal, 1981). *Naive Bayes* mengklasifikasi probabilistik berdasarkan teorema Bayes yang menggunakan asumsi yang sangat kuat pada independensi antara predictor (Torgo, 2016).

$$P(y|x) = \frac{P(x|y) \times P(y)}{P(x)} \quad (1)$$

Keterangan $P(y|x)$ adalah probabilitas kelas y (label) diberikan fitur x (data), $P(x|y)$ adalah probabilitas fitur x (data) diberikan kelas y (label), $P(y)$ adalah probabilitas prior kelas y (label), $P(x)$ adalah probabilitas prior fitur x (data).

Rumus Probabilitas Fitur:

$$P(\text{fitur} | \text{kelas}) \quad (2)$$

2.3. Algoritma Support Vector Machine

Support Vector Machine adalah teknik diskriminan, dan, karena *Support Vector Machine* memecahkan masalah optimasi cembung secara analitis, selalu mengembalikan parameter hyperplane optimal yang sama - berbeda dengan algoritma genetika (GA) atau perceptron, yang keduanya banyak digunakan untuk klasifikasi dalam pembelajaran mesin (Awad & Khanna, 2015). *Support Vector Machine* mesin pembelajaran yang meminimalkan kesalahan empiris dengan mempertimbangkan "kompleksitas" dari ruang hipotesis yang digunakan dengan juga meminimalkan norma *Reproducing Kernel Hilbert Space* (RKHS) dari solusi (Evgeniou & Pontil, 2001).

$$W \cdot X + b = 0 \quad (3)$$

Keterangan W = vektor baris berdimensi d yang mewakili arah normal ke hyperplane, X = vektor baris berdimensi d yang berhubungan dengan titik data ke- l , b = skalar atau bias.

2.4. Analisis Sentimen

Analisis sentimen adalah teknologi yang bertujuan untuk mengidentifikasi, mengeluarkan, atau memprediksi sentimen yang terdapat dalam dokumen atau teks (Liu & Zhang, 2007). Dengan menggunakan metode analisis bahasa dan teknologi pemrosesan bahasa alami untuk mengenali sentimen yang terkandung dalam dokumen atau teks tersebut. Sedangkan menurut (Otero et al., 2017), Analisis sentimen (juga disebut sebagai analisis subjektivitas atau penggalan opini atau kecerdasan buatan emosi) adalah teknik pemrosesan bahasa alami (NLP) yang mengidentifikasi pola-pola informasi dan fitur penting dari korpus teks

yang besar. Terdapat beberapa tahapan dalam analisis sentiment yaitu pengumpulan data, *preprocessing*, *sentiment lexicon*, pembobotan, *modelling*, evaluasi.

2.5. Twitter

Twitter adalah situs jejaring sosial di mana orang berkomunikasi dalam pesan singkat yang disebut *Tweet* (Ariana, 2016). Para pendidik dari seluruh dunia telah bergabung dengan *Twitter* dan menggunakannya untuk terhubung, berkomunikasi, dan berbagi satu sama lain sebagai bentuk pengembangan profesional yang berkelanjutan dan personal. Pengguna dapat menulis pesan tentang berbagai topik, berbagi informasi, dan mendiskusikan topik tertentu, serta mengungkapkan kebahagiaan dan ekspresi emosional lainnya dalam bentuk status dan *Tweet*.

3. Hasil dan Pembahasan

Tahap *data selection* melibatkan pengumpulan data melalui metode *scraping* menggunakan *Node.js* dan *Python* serta diberikan label melalui pelabelan data menggunakan *Label Encoder*. Tahap *Preprocessing* terdiri dari *case folding*, *Cleansing*, *tokenizing*, *normalization*, *Stopword removing*, dan *stemming* untuk mempersiapkan data agar siap diolah oleh algoritma analisis data. Selanjutnya, *Dataset* diubah menjadi *numpy array* pada tahap Transformation untuk kemudahan manipulasi. *Dataset* dibagi menjadi data pelatihan dan pengujian dalam rasio 70:30, 80:20, dan 90:10. Dilakukan *tf-idf vectorizing* untuk mengubah teks menjadi representasi vektor berdasarkan bobot kata. Tahap *data mining* melibatkan penerapan algoritma *Naive Bayes* dan *SVM* untuk analisis klasifikasi pada data yang telah dipreproses. Pada tahap Evaluation, akurasi model dievaluasi pada ketiga rasio pembagian data untuk memahami performa model pada setiap pengujian. Hasil dari proses *data mining* disajikan dalam bentuk *word cloud* untuk visualisasi kata-kata yang signifikan dalam *Dataset*.

3.1. Data Selection

Melakukan *scraping data* menggunakan API *Twitter* dengan kata kunci “Perkuliahan daring dan Kuliah Daring” dengan periode 01 Januari 2021 sampai 30 Mei 2023 mendapatkan 3000 data. Kemudian menyaring data dengan menghapus data *Tweets* iklan dan berita sehingga jumlah *Raw Dataset* sebesar 1260 data. Setelah itu melakukan pelabelan data manual oleh pakar bahasa dengan tujuan untuk menjaga keobjektifan data penelitian. Proses pelabelan dilakukan dengan membagi data *Tweet* menjadi dua kelas yaitu sentimen positif dan sentimen negatif. Total sentimen positif sebanyak 627 data dan sentimen negatif sebanyak 633 data. *Raw Dataset* yang sudah diberi label dinamakan *pre-Dataset*.

Tabel 1. Tabel Labelling

No.	<i>Tweet</i>	Sentiment
1.	Aku waktu semester 1 full daring karena kondisi masih pandemi. Waktu semester 2 aku mengundurkan diri karena masuk kuliah ke universitas negeri, Kak. Jadi kurang tau pasti. Tapi melihat teman teman aku di semester 2 ini klo	Positif

No.	Tweet	Sentiment
	sabtu iya kak ke kampus. Mungkin tergantung dosennya juga	
2.	Pas Selasa udah beli tiket set.6 tiba tiba jam 4 ada kuliah dadakan blm kelar kelas langsung pergi untung keburu sekarang ada lagi anjir masalahnya, seneng sih kuliahnya daring tapi w keburu udh beli tiket	Negatif
3.	Bahkan, waktu kelas 1 gue masih kuliah daring dan sebelum kelas pasti nganter adik dulu ke sekolah. Gue lebih dulu milih bangku buat dia di bangku pertama, besoknya adik gue udah di belakang karena ada satu orang tua murid yang mindahin adik gue untuk duduk di belakang.	Positif

Sumber: Hasil Penelitian (2023)

3.2. Preprocessing

Sesudah data disiapkan maka data akan masuk kedalam beberapa proses, diantaranya: *label encoder*, *case folding*, *cleansing*, *tokenizing* dan *normalization*.

- a. *Label Encoder*, pengubahan tipe data kategorik (*review* dan *sentiment*) menjadi tipe data numerik (*review* dan *target*).

Tabel 2. Hasil *Label Encoder*

No.	Tweet	Setelah Label Encoder
1.	Aku waktu semester 1 full daring karena kondisi masih pandemi. Waktu semester 2 aku mengundurkan diri karena masuk kuliah ke universitas negeri, Kak. Jadi kurang tau pasti. Tapi melihat teman teman aku di semester 2 ini klo sabtu iya kak ke kampus. Mungkin tergantung dosennya juga	1
2.	Pas Selasa udah beli tiket set.6 tiba tiba jam 4 ada kuliah dadakan blm kelar kelas langsung pergi untung keburu sekarang ada lagi anjir masalahnya, seneng sih kuliahnya daring tapi w keburu udh beli tiket	0
3.	Bahkan, waktu kelas 1 gue masih kuliah daring dan sebelum kelas pasti nganter adik dulu ke sekolah. Gue lebih dulu milih bangku buat dia di bangku pertama, besoknya adik gue udah di belakang karena ada satu orang tua murid yang mindahin adik gue untuk duduk di belakang.	1

Sumber: Hasil Penelitian (2023)

- b. *Case Folding* bertujuan untuk menyamakan karakter dalam data. Proses melibatkan perubahan seluruh huruf menjadi huruf kecil, dalam *proses case folding*, karakter-karakter 'A'-'Z' yang ada dalam data diubah menjadi karakter 'a'-'z'.

Tabel 3. Hasil *Case Folding*

No.	<i>Tweet</i>	Setelah <i>Case Folding</i>
1.	Aku waktu semester 1 full daring karena kondisi masih pandemi. Waktu semester 2 aku mengundurkan diri karena masuk kuliah ke universitas negeri, Kak. Jadi kurang tau pasti. Tapi melihat teman teman aku di semester 2 ini klo sabtu iya kak ke kampus. Mungkin tergantung dosennya juga	aku waktu semester 1 full daring karena kondisi masih pandemi. waktu semester 2 aku mengundurkan diri karena masuk kuliah ke universitas negeri, kak. jadi kurang tau pasti. tapi melihat teman-teman aku di semester 2 ini klo sabtu iya kak ke kampus. mungkin tergantung dosen nya juga.
2.	Pas Selasa udah beli tiket set.6 tiba tiba jam 4 ada kuliah dadakan blm kelar kelas langsung pergi untung keburu sekarang ada lagi anjir masalahnya, seneng sih kuliahnya daring tapi w keburu udh beli tiket.	pas selasa udah beli tiket set.6 tiba tiba jam 4 ada kuliah dadakan blm kelar kelas langsung pergi untung keburu sekarang ada lagi anjir masalah nya, seneng sih kuliah nya daring tapi gw keburu udh beli tiket
3.	Bahkan, watu kelas 1 gue masih kuliah daring dan sebelum kelas pasti nganter adik dulu ke sekolah. Gue lebih dulu milih bangku buat dia di bangku pertama, besoknya adik gue udah di belakang karena ada satu orang tua murid yang mindahin adik gue untuk duduk di belakang.	bahkan, waktu kelas 1 gue masih kuliah daring dan sebelum kelas pasti nganter adik dulu ke sekolah. gue lebih dulu milih bangku buat dia di bangku pertama, besok nya adik gue udah di belakang karena ada satu orang tua murid yang mindahin adik gue untuk duduk di belakang

Sumber: Hasil Penelitian (2023)

- c. *Cleansing* bertujuan untuk menghapus *url*, *username*, *hashtag*, tanda baca, angka dan menghapus spasi yang berlebihan.

Tabel 4. Hasil *Cleansing*

No.	<i>Tweet</i>	Setelah <i>Cleansing</i>
1.	aku waktu semester 1 full daring karena kondisi masih pandemi. waktu semester 2 aku mengundurkan diri karena masuk kuliah ke universitas negeri, kak. jadi kurang tau pasti. tapi melihat teman-teman aku di semester 2 ini klo sabtu iya kak ke kampus. mungkin tergantung dosen nya juga.	aku waktu semester full daring karena kondisi masih pandemi waktu semester aku mengundurkan diri karena masuk kuliah ke universitas negeri kak jadi kurang tau pasti tapi melihat teman teman aku di semester ini klo sabtu iya kak ke kampus mungkin tergantung dosen nya juga
2.	pas selasa udah beli tiket set.6 tiba tiba jam 4 ada kuliah dadakan blm kelar kelas langsung pergi untung keburu sekarang ada lagi anjir masalah nya, seneng sih kuliah nya daring tapi gw keburu udh beli tiket	pas selasa udah beli tiket set tiba tiba jam ada kuliah dadakan blm kelar kelas langsung pergi untung keburu sekarang ada lagi anjir masalah nya seneng sih kuliah nya daring tapi gw keburu udh beli tiket

No.	Tweet	Setelah <i>Cleansing</i>
3.	bahkan, waktu kelas 1 gue masih kuliah daring dan sebelum kelas pasti nganter adik dulu ke sekolah. gue lebih dulu milihin bangku buat dia di bangku pertama, besok nya adik gue udah di belakang karena ada satu orang tua murid yang mindahin adik gue untuk duduk di belakang	bahkan waktu kelas gue masih kuliah daring dan sebelum kelas pasti nganter adik dulu ke sekolah gue lebih dulu milihin bangku buat dia di bangku pertama besok nya adik gue udah di belakang karena ada satu orang tua murid yang mindahin adik gue untuk duduk di belakang

Sumber: Hasil Penelitian (2023)

- d. *Tokenizing* bertujuan untuk memecah kalimat menjadi kata-kata. Cara yang dilakukan pada tahap *tokenizing* adalah memotong kata pada spasi atau *whitespace*.

Tabel 5. Hasil *Tokenizing*

No.	Tweet	Setelah <i>Tokenizing</i>
1.	aku waktu semester full daring karena kondisi masih pandemi waktu semester aku mengundurkan diri karena masuk kuliah ke universitas negeri kak jadi kurang tau pasti tapi melihat teman teman aku di semester ini klo sabtu iya kak ke kampus mungkin tergantung dosen nya juga	[aku,waktu,semester,full,daring,karena,kon disi,masih,pandemi,waktu,semester,aku,m engundurkan,diri,karena,masuk,kuliah,ke,u niversitas,negeri,kak,jadi,kurang,tau,pasti,t api,melihat,teman,teman,aku,di,semester,i ni,klo,sabtu,iya,kak,ke,kampus,mungkin,ter gantung,dosennya,juga,]
2.	pas Selasa udah beli tiket set tiba tiba jam ada kuliah dadakan blm kelar kelas langsung pergi untung keburu sekarang ada lagi anjir masalah nya seneng sih kuliah nya daring tapi gw keburu udh beli tiket	[pas,selasa,udah,beli,tiket,set,tiba,tiba,jam, ada,kuliah,dadakan,blm,kelar,kelas,langsung,pergi,untung,keburu,sekarang,ada,lagi,anjirr,masalahnya,seneng,sih,kuliahnya,daring,tapi,w,keburu,udh,beli,tiket]
3.	bahkan waktu kelas gue masih kuliah daring dan sebelum kelas pasti nganter adik dulu ke sekolah gue lebih dulu milihin bangku buat dia di bangku pertama besok nya adik gue udah di belakang karena ada satu orang tua murid yang mindahin adik gue untuk duduk di belakang	[bahkan,waktu,kelas,gue,masih,kuliah,daring,dan,sebelum,kelas,pasti,nganter,adik,dulu,ke,sekolah,gue,lebih,dulu,milihin,bangku,buat,dia,di,bangku,pertama,esoknya,adik ,gue,udah,di,behalang,karena,ada,satu,orang,tua,murid,yang,mindahin,adik,gue,untuk ,duduk,di,behalang,]

Sumber: Hasil Penelitian (2023)

- e. *Normalization* bertujuan untuk konversi teks ke dalam bahasa baku sesuai dengan Kamus Besar Bahasa Indonesia (KBBI) untuk menghilangkan kata-kata gaul atau slang. Selain itu, ditambahkan pula kondisi untuk mengubah kata-kata singkatan atau inisial menjadi bentuk bahasa baku.

Tabel 6. Hasil *Normalization*

No.	Tweet	Setelah <i>Normalisasi</i>
1.	[aku,waktu,semester,full,daring,karena,kon disi,masih,pandemi,waktu,semester,aku,m	saya,waktu,semester,full,daring,karena,kon disi,masih,pandemi,waktu,semester,saya,

Analisis Sentimen Masyarakat Terhadap Perkuliahan Daring di Twitter Menggunakan Algoritma Naive Bayes dan Support Vector Machine

No.	Tweet	Setelah Normalisasi
	engundurkan,diri,karena,masuk,kuliah,ke,u niversitas,negeri,kak,jadi,kurang,tau,pasti,t api,melihat,teman,teman,aku,di,semester,i ni,klo,sabtu,iya,kak,ke,kampus,mungkin,ter gantung,dosennya,juga,]	mengundurkan,diri,karena,masuk,kuliah,ke ,universitas,negeri,kak,jadi,kurang,tahu,pas ti,tetapi,melihat,teman,teman,saya,di,seme ster,ini,kalau,sabtu,iya,kak,ke,kampus,mun gkin,tergantung,dosennya,juga
2.	[pas,selasa,udah,beli,tiket,set,tiba,tiba,jam, ada,kuliah,dadakan,blm,kelar,kelas,langs ng,pergi,untung,keburu,sekarang,ada,lagi,a njirr,masalahnya,senang,sih,kuliahnya,dari ng,tapi,w,keburu,udh,beli,tiket]	pas,selasa,sudah,beli,tiket,setengah,tiba,ti ba,jam,ada,kuliah,dadakan,belum,kelar,kel as,langsung,pergi,untung,keburu,sekarang, ada,lagi,anjirr,masalahnya,senang,sih,kulia hnya,daring,tetapi,saya,keburu,sudah,beli,ti ket.
3.	[bahkan,waktu,kelas,gue,masih,kuliah,dari ng,dan,sebelum,kelas,pasti,nganter,adik,du lu,ke,sekolah,gue,lebih,dulu,milihin,bangku, buat,dia,di,bangku,pertama,esoknya,adik,g ue,udah,di,belakang,karena,ada,satu,orang ,tua,murid,yang,mindahin,adik,gue,untuk,d uduk,di,belakang,]	bahkan,waktu,kelas,saya,masih,kuliah,dari ng,dan,sebelum,kelas,pasti,nganter,adik,da hulu,ke,sekolah,saya,lebih,dahulu,milihin,b angku,buat,dia,di,bangku,pertama,esoknya ,adik,saya,sudah,di,belakang,karena,ada,s atu,orang,tua,murid,yang,mindahin,adik,sa ya,untuk,duduk,di,belakang,

Sumber: Hasil Penelitian (2023)

- f. *Stopword Removing* Pengubahan tipe data kategorik (*review* dan *sentiment*) menjadi tipe data numerik (*review* dan *target*).

Tabel 7. Hasil *Stopword Removing*

No.	Tweet	Setelah <i>Stopword</i>
1.	saya,waktu,semester,full,daring,karena,kon disi,masih,pandemi,waktu,semester,saya,m engundurkan,diri,karena,masuk,kuliah,ke,u niversitas,negeri,kak,jadi,kurang,tahu,pasti,t etapi,melihat,teman,teman,saya,di,semeste r,ini,kalau,sabtu,iya,kak,ke,kampus,mungki n,tergantung,dosennya,juga,	[waktu,semester,full,daring,kondisi,pandem i,waktu,semester,mengundurkan,diri,masu k,kuliah,universitas,negeri,kak,jadi,kurang,t ahu,melihat,teman,teman,semester,kalau,s abtu,iya,kak,kampus,mungkin,tergantung,d osennya,]
2.	pas,selasa,sudah,beli,tiket,setengah,tiba,tib a,jam,ada,kuliah,dadakan,belum,kelar,kela s,langsung,pergi,untung,keburu,sekarang,a da,lagi,anjirr,masalahnya,senang,sih,kuliah nya,daring,tetapi,saya,keburu,sudah,beli,tik et	[pas,selasa,beli,tiket,setengah,tiba,tiba,jam ,kuliah,dadakan,kelar,kelas,langsung,pergi, untung,keburu,sekarang,anjirr,masalahnya, seneng,sih,kuliahnya,daring,keburu,beli,tik et]
3.	bahkan,waktu,kelas,saya,masih,kuliah,dari ng,dan,sebelum,kelas,pasti,nganter,adik,da hulu,ke,sekolah,saya,lebih,dahulu,milihin,b angku,buat,dia,di,bangku,pertama,esoknya, adik,saya,sudah,di,belakang,karena,ada,sa	[bahkan,waktu,kelas,kuliah,daring,kelas,ng anter,adik,sekolah,lebih,milihin,bangku,bua t,bangku,pertama,esoknya,adik,belakang,s atu,orang,tua,murid,mindahin,adik,duduk,b elakang,]

No.	Tweet	Setelah Stopword
	tu,orang,tua,murid,yang,mindahin,adik,saya ,untuk,duduk,di,belakang,	

Sumber: Hasil Penelitian (2023)

- g. *Stemming* bertujuan untuk mengubah kata-kata ke bentuk akar katanya. *Stem* (akar kata) adalah bagian dari kata yang tersisa setelah dihilangkan imbuhan (awalan dan akhiran).

Tabel 8. Hasil *Stemming*

No.	Tweet	Setelah Stemming
1.	[waktu,semester,full,daring,kondisi,pandemi,waktu,semester,mengundurkan,diri,masuk,kuliah,universitas,negei,kak,jadi,kurang,tahu,melihat,teman,teman,semester,kalau,sabtu,iya,kak,kampus,mungkin,tergantung,dosennya,]	waktu semester full daring kondisi pandemi waktu semester undur diri masuk kuliah universitas negeri kak jadi kurang tahu lihat teman teman semester kalau sabtu iya kak kampus mungkin gantung dosen
2.	[pas,selasa,beli,tiket,setengah,tiba,tiba,jam,kuliah,dadakan,kelar,kelas,langsung,pergi,untung,keburu,sekarang,anjirr,masalahnya,seneng,sih,kuliahnya,daring,keburu,beli,tiket]	pas selasa beli tiket tengah tiba tiba jam kuliah dada kelar kelas langsung pergi untung keburu sekarang anjirr masalah neng sih kuliah daring keburu beli tiket
3.	[bahkan,waktu,kelas,kuliah,daring,kelas,nganter,adik,sekolah,lebih,milihin,bangku,buat,bangku,pertama,esoknya,adik,belakang,satu,orang,tua,murid,mindahin,adik,duduk,belakang,]	bahkan waktu kelas kuliah daring kelas nganter adik sekolah lebih milihin bangku buat bangku pertama esok adik belakang satu orang tua murid mindahin adik duduk belakang

Sumber: Hasil Penelitian (2023)

3.3. Evaluation

Dataset dibagi menjadi data pelatihan dan pengujian dalam rasio 70:30, 80:20, dan 90:10. Penerapan algoritma Naive Bayes dan SVM untuk analisis akurasi model dievaluasi pada ketiga rasio pembagian data untuk memahami performa model pada setiap pengujian. Hasil yang diperoleh sebagaimana pada Tabel 10. Pada rasio 90:10 akurasi *Naive Bayes* lebih baik, pada akurasi 80:20 akurasi *Support Vector Machine* lebih baik, sedangkan pada 70:30 diperoleh hasil yang sama. Dengan demikian dengan *splitting* semakin besar kedua algoritma dapat memiliki performa yang sama.

Tabel 9. Hasil Akurasi Pada Rasio Pembagian *Dataset*

No.	Algoritma	Rasio Pembagian Splitting <i>Dataset</i>	Akurasi
1.	<i>Naive Bayes</i>	90 : 10	62%
2.	<i>Naive Bayes</i>	80 : 20	72%
3.	<i>Naive Bayes</i>	70 : 30	65%
4.	<i>Support Vector Machine</i>	90 : 10	63%
5.	<i>Support Vector Machine</i>	80 : 20	66%
6.	<i>Support Vector Machine</i>	70 : 30	65%

Sumber: Hasil Penelitian (2023)

4. Kesimpulan

Berdasarkan hasil penelitian, dua metode klasifikasi yang dipelajari, yaitu SVM dan *Naive Bayes*, menunjukkan performa yang berbeda dalam mengklasifikasikan data pada berbagai persentase rasio pembagian data. SVM menghasilkan akurasi yang beragam, dengan nilai tertinggi sebesar 63% pada rasio 90:10, 66% pada rasio 80:20, dan 65% pada rasio 70:30. Di sisi lain, *Naive Bayes* juga menunjukkan variasi akurasi, mencapai 62% pada rasio 90:10, 72% pada rasio 80:20, dan 65% pada rasio 70:30. Hasil menunjukkan bahwa rasio pembagian data 80:20 mampu menghasilkan akurasi tertinggi bagi kedua metode klasifikasi, dengan SVM mencapai 66% dan *Naive Bayes* mencapai 72%. Dengan demikian, algoritma *Naive Bayes* dapat dianggap lebih efektif dalam menyelesaikan proses pengklasifikasian pada data tersebut.

Daftar Pustaka

- Aggarwal, C. C. (1981). Data Mining: The Textbook. In *Cancer Letters* (illustrate, Issue 3). Springer, 2015. [https://doi.org/10.1016/0304-3835\(81\)90152-X](https://doi.org/10.1016/0304-3835(81)90152-X)
- Ariana, R. (2016). *A BEGINNER'S GUIDE TO TWITTER FOR EDUCATORS*. Ditch That Textbook. <https://ditchthattextbook.com/a-beginners-guide-to-Twitter-for-educators/>
- Awad, M., & Khanna, R. (2015). Efficient learning machines: Theories, concepts, and applications for engineers and system designers. *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers, April*, 1–248. <https://doi.org/10.1007/978-1-4302-5990-9>
- Damanik, A. R., Sumijan, S., & Nurcahyo, G. W. (2021). Prediksi Tingkat Kepuasan dalam Pembelajaran Daring Menggunakan Algoritma *Naive Bayes*. *Jurnal Sistim Informasi Dan Teknologi*, 3, 88–94. <https://doi.org/10.37034/jsisfotek.v3i3.49>
- Dzulfaroh, A.N., & Wedhaswary, I. D. *Hari Ini dalam Sejarah: WHO Tetapkan COVID-19 sebagai Pandemi Global*. Kompas.Com. <https://www.kompas.com/tren/read/2021/03/11/104000165/hari-ini-dalam-sejarah--who-tetapkan-COVID-19-sebagai-pandemi-global>
- Evgeniou, T., & Pontil, M. (2001). Support Vector Machines: Theory and applications. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2049 LNAI(September 2001), 249–257. https://doi.org/10.1007/3-540-44673-7_12
- Setiawan, H., & Utami, E. (2021). Analisis Sentimen *Twitter* Kuliah Online Pasca COVID-19 Menggunakan Algoritma Support Vector Machine dan Naive Bayes. *Jurnal Komtika (Komputasi Dan Informatika)*, 5(1), 43–51. <https://doi.org/10.31603/komtika.v5i1.5189>
- Liu, B., & Zhang, L. (2007). A Survey of Opinion Mining and Sentiment Analysis BT - Mining Text Data. In C. Z. Charu C. Aggarwal (Ed.), *Mining Text Data*. Springer New York, NY. <https://doi.org/https://doi.org/10.1007/978-1-4614-3223-4>
- Otero, C. E., Shaikh, S., Rossi, M., & Merino, X. (2017). Encyclopedia of Computer Science and Technology. In *Encyclopedia of Computer Science and Technology*. CRC Press.

<https://doi.org/10.1201/9781315115894-70>

Parmono. (2018). Nilai dan Norma Masyarakat. *Jurnal Filsafat*, 1–47.

Sánchez, M. (2013). *From Data Mining to Big Data & Data Science : a Computational Perspective*.

Sidik, F., Suhada, I., Anwar, A. H., & Hasan, F. N. (2022). Analisis Sentimen Terhadap Pembelajaran Daring Dengan Algoritma Naive Bayes Classifier. *Jurnal Linguistik Komputasional (JLK)*, 5(1), 34. <https://doi.org/10.26418/jlk.v5i1.79>

Torgo, L. (2016). Data Mining with R. In *Data Mining with R* (2ndEdition ed.). Chapman and Hall/CRC. <https://doi.org/10.1201/9781315399102>

Ziarko, W., & van Rijsbergen, C. J. (1994.). *RSKD '93: Proceedings of the International Workshop on Rough Sets and Knowledge Discovery: Rough Sets, Fuzzy Sets and Knowledge Discovery*. Springer-Verlag.