

ANALISIS SENTIMEN PELAKSANAAN KULIAH ONLINE MENGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE

Elia Setiana¹, Marwondo², Venia Retreva Danestiara³, Wiyanudin⁴

Universitas Informatika dan Bisnis Indonesia

^{1,3}Sistem informasi, ^{2,4} Informatika

E-mail: ¹ elia.setiana@unibi.ac.id, ² marwondo@unibi.ac.id,
veniarestreva@unibi.ac.id, ³ wiyandin19@student.unibi.ac.id

Abstrak

Tulisan ini bertujuan untuk mengetahui kepuasan mahasiswa dalam pelaksanaan kuliah secara online yang dibagi ke dalam tiga buah kelas yaitu positif, netral, dan negatif. Pengumpulan data dilakukan menggunakan twint dengan sumber dari media sosial twitter dengan jumlah data 25.000 tweet. Proses pengolahan data untuk mengetahui analisis sentimen menggunakan algoritma support vector machine. Dengan algoritma ini hasil yang didapatkan adalah 76.86% untuk tingkat akurasi dengan hasil sentimen positif. Precision didapatkan sebesar 0.49 dan recall sebesar 0.53 untuk f-1 score didapatkan sebesar 0.51.

Kata kunci : algoritma, kuliah, online

Abstract

This writing aims to assess the satisfaction of students regarding the implementation of online lectures, categorized into three classes: positive, neutral, and negative. Data collection was conducted using Twint from the social media platform Twitter, with a total of 25,000 tweets. The data processing process to determine sentiment analysis utilized the support vector machine algorithm. With this algorithm, the obtained results show an accuracy rate of 76.86% for positive sentiment. The precision is 0.49, recall is 0.53, and the F1 score is 0.51

Keywords : algorithm, lectures, online

Diajukan: 05 June 2023

Disetujui: 28 June 2023

Dipublikasi: 11 July 2023

1. PENDAHULUAN

Teknologi yang berkembang pada abad ke-21 ini merubah cara pandang manusia dalam menuangkan opini dan ekspresi, dimana mereka banyak menggunakan sosial media. Salah satu sosial media yang banyak digunakan oleh masyarakat dalam menuangkan opini dan ekspresi tersebut adalah *twitter*.

Dalam penggunaan *twitter* biasanya *user* lebih terbuka dalam berpendapat mengenai apa yang mereka pikirkan. Hal ini menjadi satu alat yang bisa digunakan dalam

menilai efektivitas selama pelaksanaan kuliah secara *online*. Sehingga *tweet* dari penilaian tersebut dapat dijadikan sebagai instrumen untuk mengukur tingkat kepuasan masyarakat mengenai pelaksanaan kuliah *online*. Agar nantinya hasil analisis sentimen tersebut dapat dijadikan sebagai bahan evaluasi mengenai perkuliahan secara *online*. Untuk mengukur tingkat kepuasan tersebut maka diperlukan analisis sentimen terhadap pelaksanaan kuliah secara *online*.

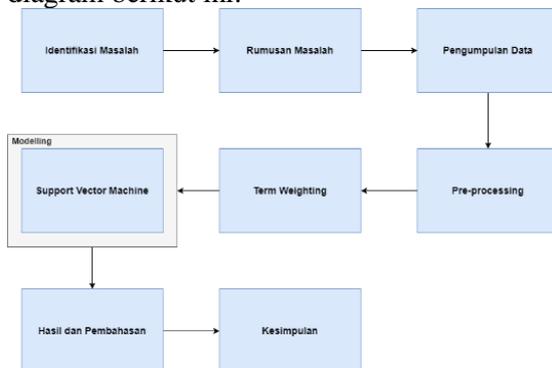
Analisis sentimen merupakan metode untuk mengukur dan menganalisis suatu

kasus atau objek tertentu. Di mana melalui analisis sentimen tersebut dapat diambil sebuah kesimpulan dan keputusan berdasarkan teks dalam bentuk kalimat ataupun dokumen. Salah satu algoritma yang dapat digunakan untuk melakukan sentimen analisis adalah *Support Vector Machine* di mana pada algoritma ini sistem akan melakukan klasifikasi melalui probabilitas kemungkinan dari data yang diperoleh.

Berdasarkan permasalahan di atas maka pada penelitian ini akan mencoba melakukan analisis sentimen pada sosial media *twitter*. Untuk mengukur opini dan pendapat masyarakat mengenai pelaksanaan kuliah *online* selama covid-19 dengan menggunakan algoritma *Support Vector Machine*. Dimana nanti akan dibagi kedalam 3 kelas yaitu positif, netral, dan negatif.

2. METODE PENELITIAN

Penelitian ini dilakukan dan dipusatkan pada studi kasus tertentu. Metode yang digunakan melalui beberapa tahap, dimana alur penelitiannya dapat dilihat melalui diagram berikut ini:



Gambar 1. Alur Penelitian

PENGUMPULAN DATA

Data yang digunakan dalam penelitian ini adalah *tweet* mengenai opini tentang perkuliahan yang diadakan secara *online*. Proses mengumpulkan data dilakukan dengan teknik *scraping* menggunakan *twint*. Data yang diambil hanya *tweet* yang mengandung kata ‘kuliah online’ dan ‘kuliah daring’.

PENGACUAN PUSTAKA

ANALISIS SENTIMEN

Analisis sentimen adalah proses menganalisa teks atau dokumen baik terstruktur ataupun tidak. Biasanya digunakan untuk teks yang bersifat online

seperti ulasan, komentar, bisa juga menggunakan media foto. Analisis ini biasa menghasilkan sebuah klasifikasi tertentu. Analisis sentimen biasanya menggunakan data yang sudah ada dengan jumlah besar.

TEKS MINING

Teks dapat diartikan sebagai proses penggalian atau penemuan informasi baru yang sebelumnya tidak diketahui. Teks *mining* mengekstrak data dari sumber – sumber yang berbeda. Teks *mining* bisa menggunakan data terstruktur ataupun data yang tidak terstruktur. Untuk data yang tidak terstruktur sebelum diproses, harus dilakukan *pre-processing data* terlebih dahulu.

PRE-PROCESSING DATA

Pre-processing data diperlukan sebelum data diklarifikasikan. yang didapat dari dataset belum bisa langsung diklasifikasikan. Karena data tersebut masih terdapat banyak simbol dan kata-kata yang tidak diperlukan. Karena itu kita memerlukan *pre-processing data* agar data lebih terstruktur dan bersih sehingga bisa diklasifikasikan. Teks *pre-processing* adalah bagian di mana data yang sudah didapat selanjutnya diolah dengan tahapan-tahapan *case folding*, *stopword removal*, *tokenization*, dan *stemming*.

TERM WEIGHTING

Term weighting adalah fitur-fitur yang sudah diekstraksi dan ditransformasi kedalam *Bag-of-Words (BoW)* yang digunakan untuk menentukan bobot pada setiap kata. *BoW* pada tahapan ini, adalah matriks yang memetakan fitur kata terhadap bobotnya sehingga bisa menghasilkan *matrix tweet* yang memiliki kolom sebanyak jumlah kata yang unik dari keseluruhan *tweet* tersebut. Pada penelitian ini pembobotan kata tersebut dihitung menggunakan teknik *TF-IDF* atau *Term Frequency - Inverse Document Frequency*.

Pembobotan *Term Frequency - Inverse Document Frequency (TF-IDF)* diperlukan untuk pembeda pada term yang ada dalam dokumen. Sehingga dalam ekstraksi bobot diperlukan untuk menentukan seberapa penting kata dalam sebuah dokumen.

Semakin besar kemunculan kata dalam dokumen maka semakin besar bobotnya.

$$w_{ij} = t_{f_{ij}} \times id_{f_{ij}}$$

$$t_{f_{ij}} = f_{(ij)}$$

$$id_{f_{ij}} = \log \frac{N}{df_{(j)}} + 1$$

Keterangan :

$$w_{ij}$$

= bobot kata t_j terhadap dokumen d_i

$$t_{f_{ij}}$$

= jumlah kemunculan term t_j dalam dokumen d_i

$id_{f_{ij}}$ = term t_j dalam koleksi dokumen

f_{ij} = frekuensi kemunculan term ke

– i pada dokumen ke – j

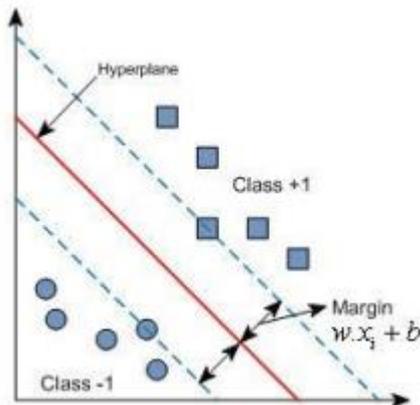
N = jumlah dokumen yang ada dalam kumpulan dokumen

$$d_{f_j}$$

= jumlah dokumen yang mengandung term t_j

SUPPORT VECTOR MACHINE

Setelah memperoleh data yang terstruktur maka proses analisis sentimen dilakukan menggunakan algoritma *Support Vector Machine*. (SVM) yaitu algoritma *machine learning* dimana algoritma tersebut menerapkan fungsi *hyperplane* pada data sehingga terbentuk bagian-bagian tiap kelas. *Hyperplane* adalah sebuah fungsi yang digunakan sebagai pemisah antar kelas yang ada. Metode *Support Vector Machine* memiliki 3 model pendekatan (kernel) yaitu kernel radial, kernel linier, dan sigmoid. Pada penelitian ini, model pendekatan yang digunakan adalah kernel linier.



Gambar 2. Struktur Support Vector Machine

Pada gambar 2 di tunjukan bahwa SVM terdiri dari dua buah kelas yaitu -1 dan +1.

Garis merah adalah *hyperplane* yang membagi data ke dalam dua kelas tersebut. Data yang mendekati *hyperplane* dibatasi oleh margin yang disebut *support vector* ditandai dengan garis putus – putus yang ditentukan oleh nilai bobot dan bias.

$$x_i w + b \geq +1 \text{ for } y_i = +1$$

$$x_i w + b \leq -1 \text{ for } y_i = -1$$

Keterangan :

w : normal bidang

b : posisi bidang relatif terhadap pusat koordinat

CONFUSION MATRIX

Confusion matrix adalah tools untuk membandingkan nilai sebenarnya dengan nilai prediksi. Dengan confusion matrix kita dapat mencari tingkat akurasi, precision, recall, dan f-1 score. Precision adalah ratio prediksi true positif dengan hasil yang ternyata memang positif. Recall adalah ratio prediksi true positif dengan data yang memang positif. F-1 score merupakan perbandingan rata-rata presisi dan recall yang dibobotkan.

PELAKSANAAN KULIAH

Pelaksanaan kuliah yang dilaksanakan oleh mahasiswa maupun dosen dapat dibagi ke dalam beberapa bagian. Pelaksanaan kuliah dilaksanakan secara offline, online, dan blend. Offline berarti pelaksanaan kuliah dilaksanakan dengan tatap muka secara langsung dilingkungan kampus. Online berarti pelaksanaan kuliah dilakukan dengan menggunakan media pendukung seperti google meet, zoom, dan yang lain. Pelaksanaan secara blend menggabungkan kuliah secara offline dan online.

3. HASIL DAN PEMBAHASAN

Data *tweet* yang dikumpulkan menggunakan *twint* sebanyak 25000 *tweet*. Data tersebut disimpan ke dalam format *csv*. Data *tweet* yang dikumpulkan hanya bahasa Indonesia dengan rentang waktu dari Januari 2021 – Juli 2022.

No.	username	tweet
1	silverlines17	@subtanyarl Iya, maka dari itu aku juga lebih senang kuliah online juga....
2	goblinahayy	Ngulang maktul dimasa begini, ga lucu bgt sih gara gara kuliah online kesiangan mulu.
3	lolipiaa	Apa tida emosi melihat jadwal kuliah berantakan kek gini:) Rumah kampus 40mntan dan aku ga kos, ya untung ini kuliah online, coba offline :)))) https://t.co/jhjmAHsoGp
4	awanbiruuuu_	overthink hari ini karena takut selama kuliah online gabisa adaptasi berujung ipk kecilðŸ©
5	meowmgie	Eh bentar lgi september yakkk, yahhh bntnr lgi mulai kuliahhh(((males bngt kuliah online, bnyk tugasnyaa:(

Tabel 1. Sample data tweet dari Twitter

Dalam proses pengumpulan data tersebut membutuhkan waktu yang cukup lama. Data berikut masih berupa data mentah yang nantinya akan dilanjutkan ke tahap *pre-processing*.

PRE-PROCESSING DATA

Tahap pertama yang dilakukan adalah *case folding*. Yaitu mengubah semua kata menjadi huruf kecil atau biasa disebut *lowercase*. Case folding digunakan untuk menjaga konsistensi data, proses ini menggunakan fungsi *lower*.

No.	Username	lowercase
1	silverlines17	@subtanyarl iya, maka dari itu aku juga lebih senang kuliah online juga....
2	Goblinahayy	ngulang maktul dimasa begini, ga lucu bgt sih gara gara kuliah online kesiangan mulu.
3	Lolipiaa	apa tida emosi melihat jadwal kuliah berantakan kek gini:) rumah kampus 40mntan dan aku ga kos, ya untung ini kuliah online, coba offline :)))) https://t.co/jhjmahsogp
4	awanbiruuuu_	overthink hari ini karena takut selama kuliah online gabisa adaptasi berujung ipk kecilðŸ©
5	Meowmgie	eh bentar lgi september yakkk, yahhh bntnr lgi mulai kuliahhh(((males bngt kuliah online, bnyk tugasnyaa:(

Tabel 2. Sample data lowercase

Tahap kedua melakukan *stopwords removal* yaitu penghapusan kata yang tidak penting beserta tanda bacanya dengan menggunakan tools *nlTK*.

No.	Username	stopword removal
1	silverlines17	iya maka juga lebih senang kuliah online juga
2	goblinahayy	ngulang masa begini lucu gara gara kuliah online kesiangan mulu.
3	lolipiaa	tida emosi melihat jadwal kuliah berantakan gini rumah kampus kos untung ini kuliah online coba offline
4	awanbiruuuu_	hari ini karena takut selama kuliah online gabisa adaptasi berujung
5	meowmgie	bentar september mulai kuliah males kuliah online tugasnya

Tabel 3. Sample data stopword removal

Tahap ketiga adalah *stemming*, yaitu mengubah kata yang ada pada tweet menjadi kata dasarnya menggunakan tools *sastrawi*.

No.	kata	kata dasar
1	temennya	temen
2	nyuruh	suruh
3	kecapean	cape
4	merasa	rasa
5	semenjak	sejak

Tabel 4. Sample data stemming

TERM WEIGHTING

Pada bagian ini adalah pembobotan pada tiap *tweet* untuk menentukan label yang nantinya diklasifikasikan menjadi netral, positif, dan negatif menggunakan *tf-idf*. Dengan jumlah data dengan kelas netral sebanyak 21064. Jumlah data dengan kelas positif sebanyak 2125 dan jumlah data dengan kelas negatif 1811.

No.	term	tweet	weighting
1	0.0	iya maka juga lebih senang kuliah online juga	
2	0.0	ulang masa begini lucu gara gara kuliah online siang mulu.	
3	0.0	tidak emosi lihat jadwal kuliah berantakan gini rumah kampus untung ini kuliah online coba offline	
4	0.0	hari ini karena takut selama kuliah online bisa adaptasi ujung	
5	-0.75	bentar september mulai kuliah males kuliah online tugas	

Tabel 5. Sample data term weighting

MODELLING DATA

Pada bagian ini data akan dibagi menjadi dua model yaitu data *training* dan data *testing*. Pada penelitian ini data *testing* digunakan sebanyak 20% dari total banyak data, yaitu sebanyak 5000 data akan digunakan sebagai data *testing*. Dan 20000 data akan digunakan sebagai data *training*.

Proses *training* ini menggunakan algoritma SVM. Dengan bantuan tool *sklearn* kita dapat dengan mudah menggunakan algoritma SVM. Cukup dengan mengisi parameter dari tool tersebut.

4. KESIMPULAN

Dari pengumpulan data ini, bagian evaluasi akan melihat berapa tingkat akurasi algoritma SVM dalam analisis sentimen. Hasil yang didapatkan algoritma SVM adalah 76.86% untuk tingkat akurasi dengan *confusion matrix* sebagai berikut.

$$\begin{bmatrix} 141 & 199 & 22 \\ 401 & 3543 & 269 \\ 56 & 210 & 159 \end{bmatrix}$$

Gambar 3. Confusion matrix

Dari *confusion matrix* tersebut maka kita dapat mengetahui *precision*, *recall* dan *f-1 score* yang didapatkan. *Precision* didapatkan sebesar 0.49 dan *recall* sebesar 0.53. Untuk *f-1 score* didapatkan sebesar 0.51.

5. SARAN

Untuk mendapatkan data yang lebih akurat, maka diperlukan penelitian dalam rentang yang lebih lebar.

REFERENCES

- [1] Arifin, O., & Sasongko, T. B. , 2018, ANALISA PERBANDINGAN TINGKAT PERFORMANSI METODE SUPPORT VECTOR MACHINE DAN NAÏVE BAYES CLASSIFIER UNTUK KLASIFIKASI JALUR MINAT SMA. *Seminar Nasional Teknologi Informasi dan Multimedia 2018*.
- [2] Fiska, R. R. (2017). Penerapan Teknik Data Mining dengan Metode Support Vector Machine (SVM) untuk Memprediksi Siswa yang Berpeluang Drop Out (Studi Kasus di SMKN 1 Sutura). *SATIN*.
- [3] Jumeilah, F. S., 2017, Penerapan Support Vector Machine (SVM) untuk Pengkategorian Penelitian. *JURNAL RESTI*, 19-25.
- [4] Mutawalli, L., Asri, M. T., & Bagye, W., 2019, KLASIFIKASI TEKS SOSIAL MEDIA TWITTER MENGGUNAKAN SUPPORT VECTOR MACHINE (Studi Kasus Penusukan Wiranto). *JIRE*.
- [5] Samsir, Ambiyar, Verawardina, U., Edi, F., & Watrianthos, R., 2020, Analisis Sentimen Pembelajaran Daring Pada *Twitter* di Masa Pandemi COVID-19 Menggunakan Metode Naive Bayes. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 157-163.
- [6] Setiawan, H., Utami, E., & Sundarmawan, 2021, Analisis Sentimen *Twitter* Kuliah Online Pasca Covid-19 Menggunakan Support Vector Machine dan Naive Bayes. *Jurnal Komtika*.
- [7] Sitanayah, V. K., Iriani, A., & Purnomo, H. D., 2020, Analisis Sentimen Transportasi Online Menggunakan Support Vector Machine Berbasis Particle Swam Optimzation. *Jurnal Nasional Teknik Elektro dan Teknologi Informasi*.