

Analisis Sentimen Kampus Merdeka Menggunakan Machine Learning

Parasian D.P Silitonga^{1*}, Irene Sri Morina², Mitra Hasibuan³, Uning Lestari⁴

^{1,3}Program Studi Teknik Informatika, Universitas Katolik Santo Thoma, Indonesia

²Rumah Sakit Umum Haji Adam Malik, Indonesia

⁴Program Studi Informatika, Institut Sains & Teknologi AKPRIND Yogyakarta, Indonesia

Email: ¹parasianirene@gmail.com, ²morina_ginting@yahoo.com, ³ mitrahasibuan01@gmail.com,

⁴uning@akprind.ac.id

INFORMASI ARTIKEL

Histori artikel:

Naskah masuk,

29 Juni 2022

Direvisi, 16 Juli 2022

Diiterima, 23 Juli 2022

Kata Kunci:

Analisis Sentimen,
Machine Learning,
Support Vector Machine,
Kampus Merdeka.

ABSTRAK

Abstract- sentiment analysis is the interpretation and classification of users' emotions (positive, negative, neutral) about a subject in text data using text analysis. Several methods can be used to perform sentiment analysis, one of which is machine learning. Machine learning is used as a tool to produce robots that can classify types of sentiment in textual data. This research was conducted to produce a machine learning model using the Support Vector Machine (SVM) method that can be used to measure the level of popularity of the independent campus program which was launched by the Ministry of Education, Culture, Research and Information Technology of the Republic of Indonesia based on data from comments or public opinion in the media. social. Based on the research conducted, it was found that the number of true positive rates was 270 records categorized as positive labels and the false positive rate was 0 records categorized as negative labels. Then the number of true negative rates is 11 records categorized as negative labels the and false negative rate is 67 records categorized as positive labels. The results of the data test found that the accuracy of the SVM algorithm was 80.75%.

Abstrak-Analisis sentimen merupakan interpretasi dan klasifikasi emosi (positif, negatif, netral) pengguna tentang suatu subjek dalam data teks dengan menggunakan analisis teks. Terdapat beberapa metode yang dapat digunakan untuk melakukan analisis sentimen, salah satunya adalah machine learning. *Machine learning* digunakan sebagai tools untuk menghasilkan robot yang mampu mengklasifikasikan jenis sentimen dalam data tekstual. Penelitian ini dilakukan untuk menghasilkan model *machine learning* dengan menggunakan metode *Support Vector Machine (SVM)* yang dapat digunakan untuk mengukur tingkat popularitas program kampus merdeka yang telah diluncurkan oleh Kementerian Pendidikan, Kebudayaan, Riset dan Teknologi Informasi Republik Indonesia berdasarkan data komentar atau opini masyarakat di media sosial. Berdasarkan penelitan yang dilakukan, ditemukan bahwa jumlah *true positive rate* adalah 270 record dikategorikan sebagai label positif dan *false positive rate* adalah 0 record dikategorikan sebagai label negatif. Kemudian jumlah *true negative rate* adalah 11 record dikategorikan sebagai label negatif dan *false negative rate* adalah 67 record dikategorikan sebagai label positif. Hasil pengujian data ditemukan bahwa tingkat akurasi algoritma SVM adalah sebesar 80,75%.

Copyright © 2019 LPPM - STMIK IKMI Cirebon open access article under the CC-BY license

Penulis Korespondensi:

Parasian D.P Silitonga

Program Studi Teknik Informatika,

Universitas Katolik Santo Thomas

Jl. Setia Budi No. 479-F Tanjung Sari, Medan, Indonesia

Email: parasianirene@gmail.com

1. Pendahuluan

Analisis sentimen merupakan interpretasi dan klasifikasi emosi (positif, negatif, netral) pengguna tentang suatu subjek dalam data teks dengan menggunakan analisis teks [1]. Dengan bantuan *analisis sentimen*, informasi yang tidak terstruktur yang dapat diubah menjadi data yang lebih terstruktur yang kemudian dapat digunakan menjelaskan opini masyarakat mengenai produk, merek, layanan, politik, atau topik lainnya. Perusahaan, pemerintah, maupun bidang lainnya kemudian memanfaatkan data-data tersebut untuk melakukan analisis *pemasaran*, umpan-balik produk, maupun layanan masyarakat [2].

Terdapat beberapa metode yang dapat digunakan untuk melakukan analisis sentimen, salah satunya adalah machine learning. Machine Learning digunakan sebagai tools untuk menghasilkan robot yang mampu mengklasifikasikan jenis sentimen dalam data tekstual [3]. Machine learning merupakan cabang dari kecerdasan buatan yang memiliki kemampuan untuk mengakses data yang ada atas perintah mereka sendiri. Machine learning mampu mempelajari data yang ada dan melakukan tugas tertentu serta mampu mempelajari algoritma dan model statistik yang diberikan.

Merdeka Belajar Kampus Merdeka (MBKM), merupakan kebijakan Menteri Pendidikan dan Kebudayaan, yang bertujuan mendorong mahasiswa untuk menguasai berbagai keilmuan yang berguna untuk memasuki dunia kerja. Kampus Merdeka memberikan kesempatan bagi mahasiswa untuk memilih mata kuliah yang akan mereka ambil [4]. Pembelajaran dalam kampus merdeka memberikan tantangan dan kesempatan untuk pengembangan kreativitas, kapasitas, kepribadian, dan kebutuhan mahasiswa, serta mengembangkan kemandirian dalam mencari dan menemukan pengetahuan melalui kenyataan dan dinamika lapangan seperti persyaratan kemampuan, permasalahan riil, interaksi sosial, kolaborasi, manajemen diri, tuntutan kinerja, target dan pencapaiannya [5].

Dengan diluncurkannya program kampus merdeka, perguruan tinggi dituntut untuk merancang dan melaksanakan program pembelajaran yang inovatif agar mahasiswa dapat meraih capaian pembelajaran mencakup aspek sikap, pengetahuan, dan keterampilan secara optimal dan selalu relevan. Dengan adanya program kampus merdeka, tidak lagi terdapat deskriminasi antara perguruan tinggi dan swasta. Semua perguruan tinggi diberikan kebebasan dan perlakuan yang sama oleh pemerintah [6].

Kebijakan merdeka belajar kampus merdeka mendapat respons bukan hanya di dalam negeri

tetapi juga di luar negeri yang tertarik dengan konsep yang dicanangkan oleh pemerintah melalui Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi. Berbagai komentar maupun opini diberikan oleh masyarakat melalui berbagai media terutama media sosial. Komentar maupun opini yang diberikan ada yang bersifat positif dan ada juga yang bersifat negatif. Komentar maupun opini terutama yang terdapat dalam media sosial menjadi sumber data yang dapat digunakan untuk mengukur tingkat popularitas sebuah program maupun produk yang diluncurkan. Dukungan maupun penolakan terhadap sebuah program dapat diukur berdasarkan komentar maupun opini publik yang terdapat dalam media sosial [7].

Berdasarkan uraian tersebut, penelitian ini dilakukan untuk menghasilkan model *machine learning* yang dapat digunakan untuk mengukur tingkat popularitas program kampus merdeka yang telah diluncurkan oleh Kementerian Pendidikan, Kebudayaan, Riset dan Teknologi Republik Indonesia berdasarkan data komentar atau opini masyarakat di media sosial.

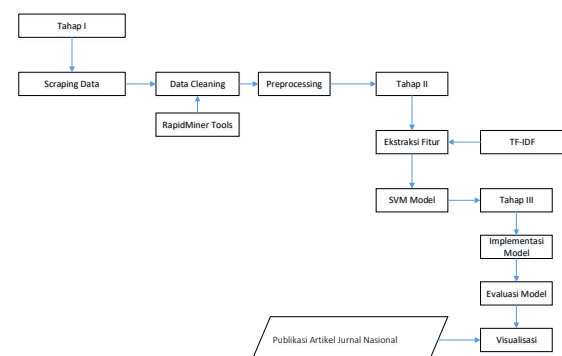
2. Metode

2.1. Datasets

Sumber data yang digunakan pada penelitian ini adalah data text hasil *crawling* dari media sosial Twitter. Data analisis sentimen diambil sebanyak 351 *tweet* dengan sentimen positif sebanyak 267 *tweet* dan sentimen negatif sebanyak 85 *tweet* yang dibagi menjadi dua jenis data dengan rasio yaitu 70% untuk data latih 30% untuk data uji.

2.2 Tahapan Penelitian

Pada penelitian ini akan dilakukan tahapan-tahapan penelitian yang dapat dilihat pada gambar 1.



Gambar 1. Tahapan Penelitian

A. Pemrosesan Data

Tahapan pemrosesan data dilakukan melalui tiga tahapan antara lain :

- 1) Data scraping; merupakan tahapan mengambil data komentar maupun opini dari media sosial *tweeter*.
- 2) *Data Cleaning*; merupakan tahapan membersihkan data dari data yang tidak diperlukan seperti *emoticon*, dan simbol-simbol yang lainnya. *Data cleaning* pada penelitian ini dilakukan dengan menggunakan *RapidMiner*.
- 3) Tahapan *preprocessing* merupakan tahapan yang dilakukan untuk menghasilkan data set. *Preprocessing* dilakukan dengan menggunakan Tokenisasi (proses memecahkan kalimat menjadi kata) *Bi-Gram* dan *Unigram*, serta *Part of Speech Tagging* yang merupakan proses pengolahan kata hasil tokenisasi ke dalam kelompok kata kerja, kata benda maupun kata sifat.

B. Pembobotan TF-IDF

Ekstraksi fitur dengan menggunakan *TF-IDF*. Ekstraksi fitur digunakan untuk menggali informasi serta merepresentasikan data sebagai vektor fitur tahapan *machine learning* selanjutnya. Ekstraksi adalah suatu proses mengubah term dari suatu teks menjadi nilai numerik yang dapat dibaca oleh komputer.

Salah satu algoritma yang dapat digunakan adalah *TF-IDF* Ekstraksi dengan fitur *TF-IDF* (*Term Frequency and Inverse Document Frequency*) adalah salah satu proses dari teknik ekstraksi fitur dengan proses memberikan nilai pada masing-masing kata yang ada pada data training. Untuk mengetahui seberapa penting sebuah kata mewakili sebuah kalimat, akan diberi nilai perhitungan. Pemberian nilai pada *TF-IDF* tergantung besarnya frekuensi kemunculan kata pada dokumen (Wang et al., 2019). Pada ekstraksi dengan fitur *TF-IDF* rumus yang digunakan untuk menghitung nilai (W) dari masing-masing dokumen terhadap kata kunci dengan formula pada persamaan 1.

$$W_{dt} = TF_{dt} * (IDF_{ft} + 1) \quad (1)$$

Dimana :

W_{dt} = nilai dokumen ke-d pada kata ke-t

TF_{dt} = jumlah kata yang dicari dalam suatu dokumen

IDF_{ft} = inverse document frequency ($\log(\frac{N}{df})$)

N = jumlah dokumen

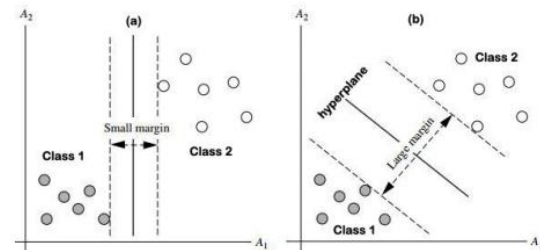
df = jumlah dokumen yang mengandung kata yang dicari

C. Support Vector Machine (SVM)

Berikutnya merupakan tahapan membangun model klasifikasi analisis sentimen menggunakan machine learning. Metode klasifikasi polaritas sentimen pada penelitian ini dilakukan dengan

menggunakan metode *Support Vector Machine (SVM)*.

SVM merupakan teknik populer yang digunakan untuk klasifikasi. Teknik *SVM* mencoba untuk menemukan pemisahan fungsi (*hyperplane*) yang paling optimum untuk memisahkan data dari kelas yang berbeda. Ilustrasi *hyperplane* di *SVM* disajikan seperti pada gambar 2.



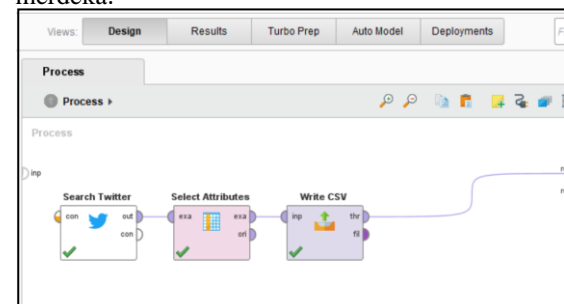
Gambar 2. Ilustrasi Hyperplane di SVM

Hasil yang diperoleh pada tahapan ini adalah model klasifikasi analisis sentimen dengan menggunakan machine learning.

3. Hasil Penelitian

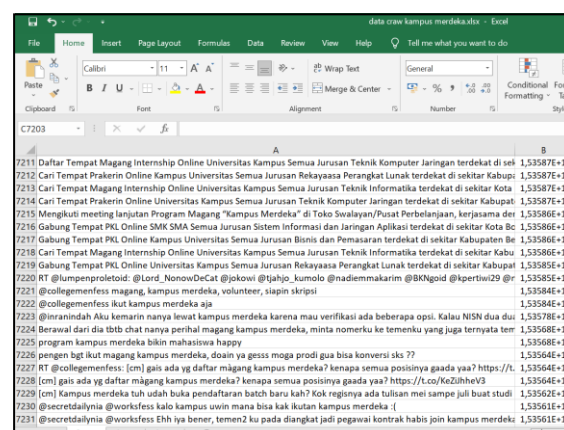
3.1. Pengambilan Data

Pada tahapan ini dilakukan pengambilan data opini dari media social twitter. Data masukan yang akan digunakan bersumber dari *tweet @kampusmerdeka*.



Gambar 3. Desain *Data Scrapping RapidMiner*

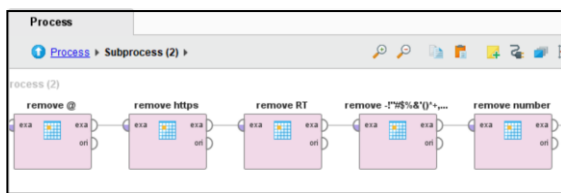
Hasil data scrapping dengan RapidMiner disimpan dalam format file excel, seperti pada Gambar 4.



Gambar 4. *Data Scrapping Result*

3.2. Data Cleaning

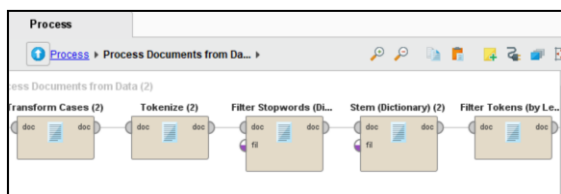
Proses berikutnya adalah tahap preparing maupun persiapan, di dalam tahap ini terdapat beberapa langkah kegiatan yaitu membangun dan membersihkan data yang nantinya akan digunakan pada tahap pemodelan data. Langkah-langkah tersebut dimulai dari tahapan penghapusan ekspresi, menghilangkan tanda at (@), hastag yang diawali dengan atribut ('#'), link yang diawali dengan atribut ('http','bit.ly') dan karakter symbol (~!@#\$%^&*()_+?<>.,?:{}[]]). Atribut yang tidak berpengaruh tersebut akan dihilangkan dari dokumen kemudian akan digantikan dengan karakter spasi.



Gambar 5. Data Cleaning

3.3. Data Preprocessing

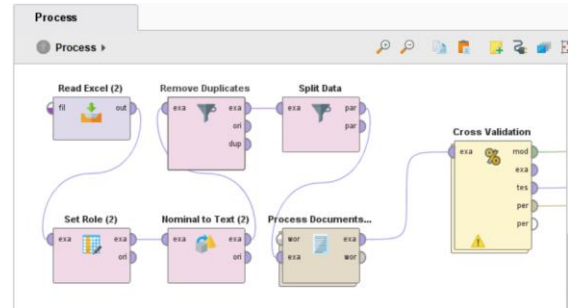
Data preparation dilakukan sebelum dataset dimasukkan kedalam model. Proses tersebut meliputi beberapa tahap, yaitu: (1) transform case untuk mengubah huruf pada text menjadi huruf kecil semua guna menyeragamkan bentuk huruf, (2) remove http untuk menghapus tautan yang terdapat di dalam tweet, (3) remove @ untuk menghapus mention di dalam tweet, (4) tokenize untuk memisahkan kata-kata pada setiap kalimat menjadi kata tersendiri sekaligus menghapus karakter yang tidak diperlukan, (5) Filter tokens by length untuk menghapus kata dengan jumlah karakter yang kurang dari nilai yang ditentukan, dalam hal ini 4 karakter, dan (6) remove stopwords untuk menghapus kata-kata yang dapat diabaikan, biasanya berupa kata keterangan dan kata sambung, dalam hal ini digunakan kamus stopwords Bahasa Indonesia.



Gambar 6. Desain Pre-Processing Data

3.4. Pemodelan

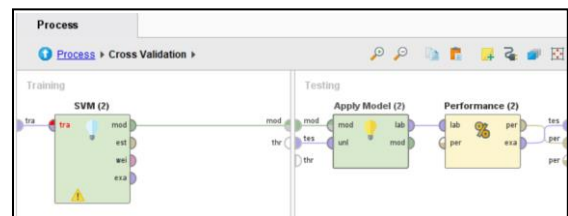
Teknik pemodelan yang digunakan dalam data mining menggunakan algoritma Support Vector Machine (SVM). Pada tahapan membangun model dilakukan dengan bantuan tools rapidminer.



Gambar 7. Desain Model Algoritma SVM

4. Pembahasan

Pengujian model parameter SVM untuk Analisis Sentimen Kampus Merdeka. Metode pengujian menggunakan Numerical Cross Validation dengan model seperti pada gambar 8.



Gambar 8. Desain Model Validation Algoritma SVM

Berdasarkan hasil pengujian model SVM didapatkan nilai akurasi yang terdapat pada Gambar menunjukkan Jumlah true positive rate adalah 270 record dikategorikan sebagai label positif dan false positive rate adalah 0 record dikategorikan sebagai label negatif. Kemudian jumlah true negative rate adalah 11 record dikategorikan sebagai label negatif dan false negative rate adalah 67 record dikategorikan sebagai label positif. Gambar 9 menunjukkan tingkat akurasi algoritma SVM adalah sebesar 80,75%.

Criterion	Table View	Plot View
accuracy	accuracy: 80.75% +/- 1.81% (micro average: 80.75%)	
recall		
AUC (optimal)		
AUC		
AUC (perclass)		
	true positive	false negative
pred positive	270	67
pred negative	0	11
class recall	100.00%	14.19%
	class precision	
	85.12%	100.00%

Gambar 9. Nilai Accuracy Algoritma SVM

Pengukuran akurasi menggunakan rumus sebagai berikut:

$$\begin{aligned}
 \text{Accuracy} &= \frac{TP+TN}{TP+TN+FP+FN} * 100 \\
 &= \frac{270+11}{270+11+67+0} * 100 \\
 &= \frac{281}{348} * 100 \\
 &= 80,75\%
 \end{aligned}$$

5. Kesimpulan

Berdasarkan hasil penelitian yang diperoleh menunjukkan bahwa analisis sentimen yang

dilakukan berdasarkan data sentimen yang diperoleh dari media sosial Twitter menghasilkan sentiment positif terhadap program kampus merdeka, dengan jumlah positif rate sebanyak 270 record dan true negative rate sebanyak 11 record. Tingkat akurasi model svm yang dibangun pada penelitian ini adalah sebesar 80,75%.

Ucapan Terima kasih

Terima kasih disampaikan kepada Universitas Katolik Santo Thomas, yang telah mendanai penelitian ini. Terima kasih juga disampaikan kepada teman-teman yang ada di Fakultas Ilmu Komputer Universitas Katolik Santo Thomas, yang telah memberikan banyak bantuan bagi peneliti sehingga penelitian ini dapat diselesaikan dengan baik.

Daftar Pustaka

- [1] M. A. Sghaier and M. Zrigui, "Sentiment analysis for Arabic e-commerce websites," 2016. doi: 10.1109/ICEMIS.2016.7745323.
- [2] S. Bandari and V. V Bulusu, "Survey on Ontology-Based Sentiment Analysis of Customer Reviews for Products and Services," *Advances in Intelligent Systems and Computing*, vol. 1079. Computer Science and Engineering, JNTU Hyderabad, Hyderabad, India, pp. 91–101, 2020. doi: 10.1007/978-981-15-1097-7_8.
- [3] J. R. Saura, A. Reyes-Menendez, and P. Palos-Sanchez, "A feeling analysis in Twitter with machine learning: Capturing sentiment from #BlackFriday offers ," *Espacios*, vol. 39, no. 42, 2018, [Online]. Available: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85055057928&partnerID=40&md5=830513edad8b62fe8adf0de560d53a07>
- [4] *Buku Panduan Merdeka Belajar - Kamus Merdeka*. Direktorat Jenderal Pendidikan Tinggi Kementerian Pendidikan dan Kebudayaan, 2020.
- [5] *Buku Pegangan Dosen Pembimbing Lapangan (DPL)*. Direktorat Jenderal Pendidikan Tinggi Kementerian Pendidikan dan Kebudayaan, 2021.
- [6] "Program Kampus Merdeka Ajak Mahasiswa Indonesia menjadi SDM Kreatif dan Adaptif – Direktorat Jenderal Pendidikan Tinggi Kementerian Pendidikan dan Kebudayaan Republik Indonesia." <https://dikti.kemdikbud.go.id/kabar-dikti/kabar/program-kampus-merdeka-ajak-mahasiswa-indonesia-menjadi-sdm-kreatif-dan-adaptif/> (accessed Aug. 23, 2021).
- [7] W. Kaswidjanti, H. Himawan, and P. D. P. Silitonga, "The accuracy comparison of social media sentiment analysis using lexicon based and support vector machine on souvenir recommendations," *Test Eng. Manag.*, vol. 82, no. 3–4, pp. 3953–3961, 2020.