

Deteksi Emosi Berbasis Teks Untuk Menganalisis Kuliah Daring Selama Masa Pandemi Menggunakan Algoritme *Support Vector Machine*

Text Based Emotion Detection To Analyze Online Lecture During Pandemic Using Support Vector Machine Algorithm

1st Damas Adiyanto
Fakultas Teknik Elektro
Universitas Telkom
Bandung, Indonesia
damasadiyanto@student.telkomuniversity.ac.id

2nd Casi Setianingsih
Fakultas Teknik Elektro
Universitas Telkom
Bandung, Indonesia
setiacasie@telkomuniversity.ac.id

3rd Fussy Mentari Dirgantara
Fakultas Teknik Elektro
Universitas Telkom
Bandung, Indonesia
fussymentari@telkomuniversity.ac.id

Abstrak—Pandemi Covid-19 di Indonesia semakin menyebar luas, banyak mahasiswa yang terpaksa harus melakukan aktivitas belajar melalui daring, kuliah daring ini memiliki berbagai macam efek pada mahasiswa, ada yang senang, sedih, takut, cinta, dan marah, hal ini menyebabkan mahasiswa sering menunjukkan emosinya melalui media sosial salah satunya Twitter. Pada Tugas Akhir ini digunakan algoritma Support Vector Machine sebagai metode klasifikasi teks berbahasa Indonesia. Berdasarkan penelitian yang dilakukan oleh Shaver, terdapat lima kategori emosi dasar bahasa Indonesia yaitu marah, senang, sedih, takut, dan cinta. Pembagian data dibagi menjadi data tiga label emosi marah, senang, dan cinta, dan 4 label emosi marah, senang, takut, cinta. Data yang digunakan diambil dari scraping data twitter dan data Github. Hasil penelitian Tugas Akhir ini, menunjukkan bahwa sistem algoritma Support Vector Machine sebagai pendeteksi emosi berbasis teks yang berkaitan dengan kuliah online, mendapatkan hasil terbaik pada pengujian tiga emosi mendapatkan hasil akurasi sebesar 81.81% pada partisi data 0.3. Sementara dengan empat emosi mendapatkan akurasi sebesar 71.86% pada partisi data 0.2. Sementara pada lima

emosi mendapatkan akurasi sebesar 60.62% pada partisi data 0.3.

Kata kunci—deteksi emosi, kuliah online, support vector machine, twitter

Abstract—The Covid-19 pandemic in Indonesia is increasingly spreading, many students are forced to do online learning activities, these online lectures have various effects on students, some are happy, sad, afraid, love, and angry, this causes students to often show emotions through social media, one of which is Twitter. In this final project, the Support Vector Machine algorithm is used as a method of classifying Indonesian text. Based on research conducted by Shaver, there are five basic categories of Indonesian emotions, namely anger, joy, sadness, fear, and love. The division of the data is divided into three emotional labels angry, happy, and love, and 4 emotional labels angry, happy, afraid, love. The data used is taken from twitter data scraping and Github data. The results of this Final Project, show that the Support Vector Machine algorithm system as a text-based emotion detector related to online lectures, gets the best results on the three-emotion test, getting an accuracy of 81.81% on a data partition of 0.3. Meanwhile, the four

emotions get an accuracy of 71.86% on the 0.2 data partition. While the five emotions get an accuracy of 60.62% on a data partition of 0.3.

Keywords—emotion detection, online lecture, support vector machine, twitter

I. PENDAHULUAN

Pandemi covid-19 di Indonesia semakin memburuk dan banyak mahasiswa yang terpaksa harus melakukan aktivitas belajar melalui daring, kuliah daring ini memiliki berbagai macam efek pada mahasiswa, ada yang senang, sedih, takut, cinta, dan amarah, dengan ini mahasiswa sering kali menunjukkan emosinya melalui media sosial salah satunya Twitter.

Respon yang diberikan sangat bervariasi, ada yang merasa senang dan merasa diuntungkan ada juga yang merasa tidak diuntungkan atau terbebani oleh kebijakan pemerintah terhadap kuliah daring. Dari hal tersebut akan dilakukan sebuah penelitian untuk mendeteksi emosi berbasis teks menggunakan machine learning.

Dalam mendeteksi emosi dari Twitter telah digunakan salah satu metode machine learning yang bernama Support Vector Machine. Support Vector Machine biasa digunakan untuk klasifikasi teks, cara kerjanya mencari Hyperplane dan mencari jarak antara satu kelas dengan kelas lainnya.

Penelitian ini dilakukan dengan tujuan untuk mengetahui emosi seseorang pada media teks, yang mana dapat menjadi masukan bagi instansi pendidikan dalam meningkatkan layanan dan mutu Pendidikan di Indonesia.

Penelitian mengenai deteksi emosi berbasis teks ini menggunakan metode Support Vector Machine, penelitian ini sudah pernah dilakukan sebelumnya dan hasilnya berupa tabel emosi, persentase akurasi emosi. Tetapi dari penelitian itu belum diimplementasikan untuk memvisualisasikan datanya dengan berbasis web dan masih berbahasa inggris. Maka dari itu, pada tugas akhir ini, telah dibuat deteksi emosi berbasis text dimana data diambil dari twitter yang membahas mengenai kegiatan kuliah online di Indonesia menggunakan metode Support Vector Machine yang visualisasi hasil analisis akan ditampilkan dalam bentuk web.

II. KAJIAN TEORI

A. Emosi

Emosi adalah keadaan psikologis dimana mengandung tiga komponen berbeda yaitu pengalaman subjektif, respons fisiologis, dan komponen ekspresif atau perilaku [1]. Emosi juga didefinisikan sebagai reaksi tubuh terhadap situasi tertentu. Biasanya berkaitan dengan aktivitas berpikir (kognitif) seseorang, yaitu jenis dan intensitas emosi yang timbul dari persepsi terhadap situasi yang sedang terjadi. Emosi merupakan aspek yang memiliki pengaruh kuat terhadap sikap manusia [2]. Dalam tugas akhir ini, emosi dibagi menjadi 5, yaitu Fear, Happy, Sadness, Love dan angry.

B. Web Scraping

Web Scraping adalah proses pengumpulan data dari web dengan topik tertentu sesuai dengan pengguna yang dilakukan secara sistematis [3]. Kebanyakan kasus web scraping adalah permintaan otomatis ke web server untuk meminta data dan kemudian parsing data tersebut untuk mendapatkan informasi yang diinginkan [4].

C. Preprocessing text

Preprocessing text merupakan Teknik untuk menghilangkan text atau paragraph yang tidak diinginkan sehingga teks yang diinginkan bisa di seleksi supaya menjadi lebih terstruktur, Preprocessing Text memiliki beberapa tahapan yang dilakukan dalam Preprocessing Text. Pada tahap ini akan dilakukan proses pemilihan data pada setiap dokumen. Proses ini terdiri dari beberapa proses pembersihan dokumen seperti tokenizing, case folding, dan stemming [5].

D. Pembobotan TF-IDF

TFIDF merupakan sebuah metode yang menggunakan integrasi antara term frequency, dengan inverse term frequency (TF). Term frequency dihitung dengan *term frequency* ke-i menjadi frekuensi kemunculan term ke-i dalam dokumen ke-j. *Inverse Document Frequency* (IDF) Adalah logaritma dari perbandingan jumlah total dokumen dalam korpus dengan jumlah dokumen yang secara matematis menggambarkan istilah Nilai yang didapatkan dengan mengalikan keduanya [6]. Berikut adalah rumus pembobotan TF-IDF

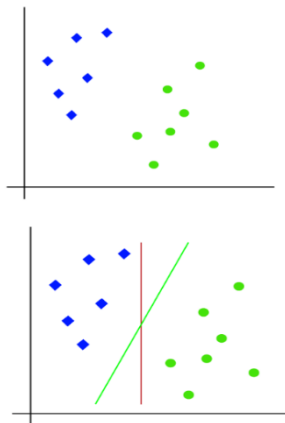
$$tf_i = \frac{freq_i(d_j)}{\sum_{i=1}^k freq_i(d_j)}$$

$$idf_1 = 1 + \log\left(\frac{N + 1}{df(t) + 1}\right)$$

$$(tf - idf)_{if} = tf_i(d_j) * idf_i$$

E. Support Vector Machine

Support Vector Machine (SVM) dikembangkan oleh Boser, Guyon, dan Vapnik dan pertama kali dipresentasikan pada Lokakarya Tahunan 1992 tentang Teori Pembelajaran Komputasi. Konsep dasar metode SVM sebenarnya merupakan gabungan atau kombinasi dari teori-teori komputasi yang ada pada tahun sebelumnya, seperti batas hyperplane. Terminologi SVM menggambarkan jarak atau margin antar kategori [7].



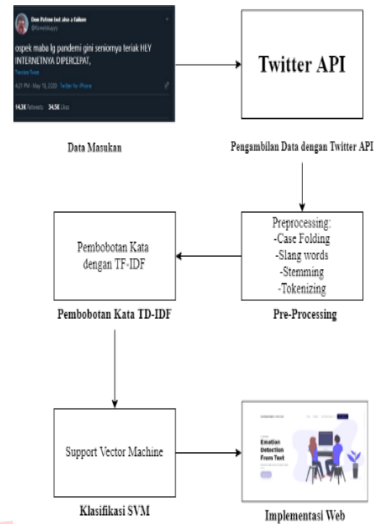
GAMBAR 1. SVM

III. METODE

A. Desain Sistem

Dalam Perancangan sistem klasifikasi emosi berdasarkan teks yang mengandung kuliah online berbahasa Indonesia yang diambil dari tweet pada Twitter dengan menggunakan algoritma *Support Vector Machine* dan kemudian akan diimplementasikan pada *website*.

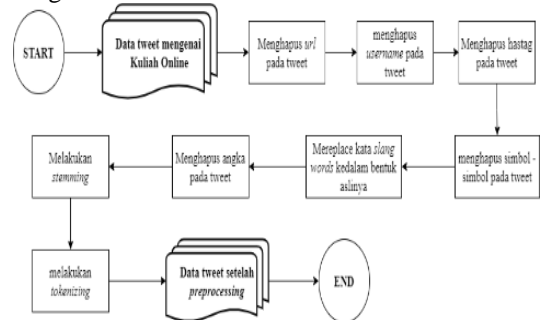
Proses yang dilakukan pada sistem ini terdiri dari pengambilan *tweet* dari Twitter. Pengambilan *tweet* atau data dari Twitter dilakukan dengan *web scraping* dengan menggunakan *library Tweepy*



GAMBAR 2. Desain

untuk pengambilan data. Kemudian data yang didapat akan dilakukan preprocessing dan kemudian diberi pelabelan emosi secara manual oleh ahli bahasa yang dikategorikan berdasarkan pelabelan emosi Shaver. Setelah itu data yang didapat telah diberi label oleh ahli bahasa, data akan di training dan testing terhadap data tersebut.

Setelah model sudah dibuat, model akan diimplementasikan di aplikasi berbasis website yang akan memberikan informasi mengenai emosi mahasiswa selama kuliah online pada masa pandemi covid-19 dimana data diambil dengan cara scraping menggunakan library Tweepy dan juga diambil melalui Github. Pada website ini pengguna dapat melihat dataset yang digunakan pada penelitian ini, serta dapat mengetahui informasi terkait algoritma yang digunakan serta akurasi yang dihasilkan dari tiap algoritma. Pengguna juga dapat mencari tweets berdasarkan kata kunci yang diinginkan.



GAMBAR 3. Flowchart Preprocessing

B. Pre-Processing

Ada beberapa tahap yang dilalui dalam tahap pre-processing, pada sistem ini sebelum di implementasikan ke website. Tahapan yang dimaksud dapat dilihat pada gambar 3 dibawah ini.

1. Case Folding

Proses ini dilakukan untuk membersihkan teks dari angka, hastag, url, dan tanda baca, karena hal tersebut tidak berpengaruh pada proses klasifikasi, dan akan dilakukan merubah kata *slang words* kedalam bentuk kata yang baku. Hal ini dilakukan supaya kata slang yang ada didalam teks menjadi kata baku. Setelah mengubah kata slang menjadi bahasa baku selanjutnya adalah merubah semua kalimat menjadi huruf kecil.

2. Stemming

Stemming merupakan suatu proses transformasi kata-kata yang terdapat dalam suatu dokumen ke kata-kata akarnya dengan aturan-aturan tertentu, atau dapat dianggap sebagai pembuangan imbuhan pada kata.

3. Tokenizing

Tokenizing adalah proses untuk memecah kalimat dokumen supaya menjadi kata-kata yang menyusunya.

C. TF-IDF

Pembobotan teks yang digunakan pada penelitian ini adalah TF-IDF (Term Frequency-Inverse Document Frequency). Pada TF-IDF pembobotan nilai pada kata berdasarkan banyaknya frekuensi munculnya suatu kata pada suatu dokumen.

Pada metode ini perhitungan untuk mencari nilai TF yang nantinya akan berfungsi untuk mengetahui banyaknya suatu kata didalam suatu dokumen. Jika ada kata ada di suatu dokumen maka akan diberi nilai 1. Namun jika tidak ada didalam dokumen maka akan diberi nilai 0. Berikut langkah-langkah pembobotan TF-IDF.

1. Memecah semua kalimat tweet menjadi kata yang berfungsi untuk mengetahui kata-kata yang muncul.
2. Menghitung jumlah kata yang terdapat pada setiap tweet.
3. Menghitung banyaknya kata dari suatu tweet menggunakan rumus 1.

4. Menghitung IDF dengan rumus 2 pada setiap kata yang muncul

5. Menghitung TFIDF dari setiap kata yang muncul pada sebuah tweet menggunakan rumus 3.

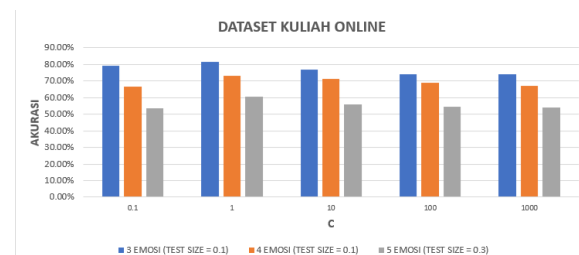
D. Support Vector Machine

Algoritma klasifikasi yang digunakan pada tugas akhir ini adalah Support Vector Machine. Tujuan dari metode ini adalah mengklasifikasikan data dengan membandingkan data training dan testing dengan nilai C tertentu.

IV. HASIL DAN PEMBAHASAN

A. Pengujian Nilai C Support Vector Machine

Pada pengujian ini adalah untuk mencari akurasi terbaik dari *Support Vector Machine* dengan cara menggunakan kernel linear dan mengubah nilai C. nilai C yang digunakan dalam pengujian ini bernilai 1000, 100, 10, 1, 0.1.



GAMBAR 4. Hasil Pengujian Dataset Kuliah Online

Berdasarkan hasil pengujian pada gambar 4 yang dilakukan menggunakan dataset 5 label emosi sebanyak 8000 data, nilai C terbaik, dataset kuliah online menghasilkan akurasi sebesar 60.62%, 4 label emosi sebanyak 6821 data, dataset kuliah online menghasilkan akurasi sebesar 73.35%, 3 label emosi sebanyak 5170 data, dataset kuliah online menghasilkan akurasi sebesar 81.81%.

B. Pengujian Confusion Matrix

Setelah mencari nilai C yang berbeda, pengujian kinerja sistem yang selanjutnya yaitu melakukan Confusion Matrix, pada Confusion Matrix terdapat nilai True Positive, false positive, true negative, false negative. Nilai tersebut nantinya akan digunakan untuk melihat keberhasilan sistem

berdasar besar akurasi, presisi dan *recall*, hasil confusion matrix terdapat pada tabel 1.

TABEL 1. Hasil Confussion Matrix 3 Emosi Kuliah Online

	Label Kelas Hasil Klasifikasi			Total
	Marah	Senang	Cinta	
Marah	529	86	7	622
Senang	65	430	47	542
Cinta	17	60	310	387
Total	611	576	364	1551
Akurasi	81.81%			
Presisi	82.33%			

	Label Kelas Hasil Klasifikasi			Total
	Marah	Senang	Cinta	
<i>Recall</i>	81.33%			

C. Pengujian dengan Menggunakan K-Fold Cross Validation

Pengujian Kinerja dari algoritma *Support Vector Machine* dengan mengubah nilai C. untuk mengetahui nilai C tersebut merupakan akurasi terbaik diperlukan proses cross validation dengan membagi $cv = 10$ atau data tersebut dibagi menjadi 10 dengan membagi rata data menjadi data training dan testing. Pengujian K-Fold bisa dilihat pada tabel 2 berikut.

TABEL 2. Hasil K-Fold Cross Validation Data Kuliah Online

Label/C	CV										Mean	Akurasi
	1	2	3	4	5	6	7	8	9	10		
3 /1	81%	79%	80%	84%	79%	82%	80%	82%	82%	83%	81%	81.81%
4 /1	72%	72%	70%	73%	71%	73%	72%	72%	72%	69%	72%	73.35%
5 /1	62%	59%	59%	60%	59%	55%	57%	60%	63%	60%	59%	60.24%

Pada tabel 2 bisa dilihat rata-rata akurasi yang didapat dengan menggunakan pengujian *K-Fold Cross Validation* dengan nilai $CV=10$ dan hasilnya tidak terlalu jauh dari akurasi yang diperoleh dari hasil menggunakan algoritma *Support Vector Machine*. Hal ini bisa disimpulkan bahwa algoritma yang digunakan sudah berjalan dengan baik.

V. KESIMPULAN

A. Kesimpulan

Berdasarkan dari hasil Tugas Akhir ini, dapat ditarik beberapa kesimpulan yaitu:

1. Performa terbaik didapatkan pada perandingan data *training* 90% dan data *testing* 10%
2. Nilai C terbaik pada algoritma *Support Vector Machine*

adalah 1 dan mendapatkan akurasi sebesar 81.81%

3. Hasil dari akurasi yang didapat tidak berbeda jauh dengan hasil dengan hasil *K-Fold Cross Validation*

B. Saran

Adapun saran dari penulis untuk penelitian yang akan datang

1. Untuk meningkatkan akurasi bisa dilakukan dengan menggunakan algoritma *deep learning*.
2. Untuk penelitian selanjutnya, perlu dilakukan pembersihan dan perubahan bahasa daerah seperti Bahasa Sunda, Bahasa Jawa, dll kedalam Bahasa Indonesia yang terambil saat proses pengambilan data. Selain itu diperlukan pencarian data tweet yang

lebih dalam untuk kelas takut dan sedih.

REFERENSI

- [1] P. Ekman, "Handbook of Cognition and Emotion," in *Basic Emotion*, 1999, pp. 45-60.
- [2] D. B. Lindsley, *Handbook of experimental psychology*, John Wiley & Sons, 2018.
- [3] D. Glez-Peña, A. Lourenco, H. López-Fernández and M. Reboiro-Jato, "Web scraping technologies in an API world," *Briefing in Bioinformatics*, pp. 1-10, 2013.
- [4] R. Mitchell, *Web Scraping with Python: Collecting Data from the Modern Web*, O'Reilly Media, 2015.
- [5] C. D. Manning, P. Raghavan and H. Schütze, *An Introduction to Information Retrieval*, Cambridge: Cambridge University Press, 2009.
- [6] K. S. JONES, "INDEX TERM WEIGHTNING," *Computer Laboratory*, , pp. 4-5, 1973.
- [7] D. Richard O, H. Peter E and S. David G, *Pattern Classification and Scene Analysis*, California: Ricoh California Research Center, 1995.