Deteksi Emosi Berbasis Teks Untuk Menganalisis Kuliah Daring Selama Masa Pandemi Menggunakan Algoritme *Naive Bayes*

Text Based Emotion Detection For Analysis Online Lecture During Pandemic Using Naive Bayes Algorithm

1st Dian Rezky Wulandari
Fakultas Teknik Elektro
Universitas Telkom
Bandung, Indonesia
dianrezx@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 muncul pertama kali pada tahun 2020 di Indonesia. Hal ini menyebabkan perubahan di berbagai aspek salah satunya pendidikan dan sosial. Media sosial banyak digunakan untuk mengekspresikan banyak hal seperti mengungkapkan emosi mereka terkait kuliah daring seperti pada Twitter. Penelitian ini bertujuan untuk mengetahui bagaimana emosi yang timbul terkait kuliah daring. Pada Tugas Akhir ini digunakan algoritme Naïve Bayes. Berdasarkan penelitian yang dilakukan oleh Shaver, emosi tersebut terdiri dari lima kelas yaitu senang, sedih, marah, takut, dan cinta. Pembagian data dibagi menjadi empat kelas emosi (senang, marah, takut, dan cinta) dan tiga kelas emosi (marah, senang, dan cinta). Data yang digunakan yaitu tweets yang diambil pada Github dan juga dilakukan web scraping pada Twitter yang diimplementasikan kedalam website. Pemodelan yang telah dibuat dilakukan pengujian menggunakan metode Confusion Matrix untuk mengetahui apakah model dapat dikatakan baik. Hasil menunjukkan bahwa sistem pendeteksi emosi berbasis teks mendapatkan akurasi sebesar 80% untuk tiga kelas emosi dengan test size 0.5, 73.20% untuk empat kelas emosi dengan test size 0.1, dan 60.87% untuk lima kelas emosi dengan test size 0.1.

Kata Kunci: Deteksi Emosi, Naïve Bayes, Text Processing.

Abstract*—*The Covid – 19 pandemics first appeared on 2020 in Indonesia. This causes changes in various aspects such as education and social life. Social media is widely used to express many things such as expressing their emotions regarding online lectures like on Twitter. In this final project, Naïve Bayes algorithm is used. Based on research conducted by Shaver, these emotions will be divided into five classes like happy, sad, angry, fear, and love. The data is divided into four emotion classes (happy, angry, fear, and love) and three emotion classes (angry, happy, and love). The data used are tweets taken on Github and also web scraping on Twitter which is implemented into the website. The model that has been made is tested using Confusion Matrix to find out whether the model can be said to be good. The results show that the show that emotion detection system get 80% of accuracy for three class of emotions with test size 0.5, 73.20% for four class of emotion with test size 0.1, and 60.87% for five class of emotions with test size 0.1.

Keywords: Emotion Detection, Naïve Bayes, Text Processing.

I. PENDAHULUAN

Pandemi Covid–19 pertama kali muncul di Wuhan, China. Covid–19 merupakan penyakit baru

yang menular disebabkan oleh corona virus versi baru yaitu Sars–CoV–2 yang menyebabkan gangguan ringan pada sistem pernapasan, infeksi paru – paru yang berat, dan bahkan dapat menyebabkan kematian. Virus Covid–19 kemudian mulai menyebar keseluruh dunia dan masuk di Indonesia pertama kali pada 2 Maret 2020 [1]. Hal ini menyebabkan perubahan di berbagai aspek salah satunya pendidikan dan sosial sehingga membuat penggunaan internet di Indonesia semakin meningkat [2]. Media sosial banyak digunakan untuk mengekspresikan banyak hal mulai dari aktivitas sehari – hari, mengungkapkan keluh kesah seperti tentang kuliah online, dan bahkan banyak yang melampiaskan emosi mereka dalam bentuk teks seperti pada Twitter.

Emosi adalah salah satu aspek penting yang tidak dapat dipisahkan dari suatu individu sejak lahir dan akan terus berkembang sesuai dengan lingkungannya. Emosi mewakili suatu perasaan dan pikiran unik, keadaan biologis dan psikologis dan suatu kecenderungan untuk berperilaku [3]. Seiring dengan meningkatnya penggunaan internet khususnya media sosial menyebabkan jumlah data teks yang tersedia semakin banyak dan tentunya data tersebut menyimpan informasi seperti sentimen dan emosi. Volume data yang bersifat banyak membuat analisis terhadap emosi atau sentimen menjadi sangat sulit. Namun hal ini dapat diatasi dengan menggunakan kecerdasan buatan.

Penelitian mengenai deteksi emosi berbasis teks menggunakan algoritme Naïve Bayes sudah banyak dilakukan seperti contoh oleh Hema Krishnan, dkk [4]. Penelitian terkait deteksi emosi berbasis teks bahasa Indonesia belum banyak diteliti. Penelitian terkait emosi dari siswa atau mahasiswa mengenai kuliah online yang dilaksanakan di Indonesia selama masa pandemi sudah pernah dilakukan namun sampel yang diambil hanya di Pulau Madura dan dilakukan dengan mengisi questionaire [5]. Maka dari itu, pada Tugas Akhir ini dilakukan deteksi emosi berbasis text dimana data akan diambil dengan web scraping dari Twitter dan juga Github yang membahas mengenai kegiatan kuliah online di Indonesia. Deteksi emosi ini menggunakan algoritme Naïve Bayes yang diimplementasikan dalam bentuk website.

II. KAJIAN TEORI

A. *Web Scraping*

Web scraping adalah proses pengumpulan data dari web dengan topik tertentu sesuai dengan pengguna yang dilakukan secara sistematis [6]. Pada praktiknya, web scraping mencakup berbagai teknik dan teknologi pemograman seperti analisis data dan keamanan informasi [7]. Tweepy adalah library dari Python untuk mengakses API Twitter. Tweepy adalah library dari Python untuk mengakses API Twitter. Tweepy adalah library dari Python untuk mengakses API Twitter. Tweepy adalah library dari Python untuk mengakses API Twitter. Twitter adalah situs microblogging yang memungkinkan user untuk mengirim pesan teks hingga 140 karakter yang disebut tweet [8]. pada Februari tahun 2010, pengguna Twitter mengirimkan 50 juta tweets setiap harinya [9]. Banyak yang dapat dilakukan pengguna pada aplikasi Twitter seperti pengguna dapat berinteraksi satu sama lain ataupun dapat mengikuti status orang lain dengan cara mengikuti pengguna terkait [10]. Twitter API merupakan suatu aplikasi yang dapat digunakan untuk mengambil dan menganalisis data Twitter secara ter-program dan juga membangun percakapan di Twitter [11].

B. Emosi Pada Manusia

Emosi adalah keadaan psikologis dimana mengandung tiga komponen berbeda yaitu pengalaman subjektif, respons fisiologis, dan komponen ekspresif atau perilaku [12]. Contoh model yang umum digunakan adalah model Robert Plutchick, model Paul Ekman, dan model OCC. Pada model Paul Ekman, emosi dibagi menjadi enam kelas dasar yaitu sedih, bahagia, jijik, marah, terkejut, dan takut [12]. Kebanyakan dari klasifikasi emosi digunakan untuk deteksi emosi teks berbahasa Inggris seperti klasifikasi emosi menurut Robert Plutchick [13]. Pada akhirnya, Shaver mendefinisikan 5 kelas emosi yang digunakan untuk leksikon berbahasa Indonesia yaitu senang, sedih, marah, takut, dan cinta [14].

C. Text Processing

Text preprocessing merupakan tahap pertama dalam klasifikasi teks yang mengubah teks menjadi sebuah data yang akan diproses lebih lanjut. Pada tahap ini akan dilakukan proses pemilihan data pada setiap dokumen. Proses ini terdiri dari beberapa proses pembersihan dokumen seperti case folding, merubah slang words, stemming, dan tokenizing [15].

1. *Case Folding* dan *Slang Words*

Case folding adalah proses preprocessing kata yang dilakukan untuk integrasi karakter dalam data seperti merubah huruf kapital ke huruf kecil atau menghapus delimiter [16]. Delimiter adalah urutan satu atau lebih karakter yang digunakan untuk mendefinisikan pembatas [17]. Slang words adalah bahasa gaul yang sering digunakan dikalangkan remaja atau anak muda baik di Amerika, Inggris, atau bahkan di Indonesia [18]. Penelitian terkait slang word pada bahasa Indonesia telah dilakukan oleh N. Aliyah Salsabila, dkk [19]. Contoh dari case folding dan slang words dapat dilihat pada Tabel 1.

TABEL 1. Proses *Case Folding* dan Merubah *Slang Words*

Data Input	Data Output
Kuliah online membawa berkah akhirnya dapat juga	kuliah daring membawa berkah akhirnya dapat juga beasiswa

beasiswa Ã‚Â¥Ã‚Âº Ã¢â€žÂ¢Ã‚Â•	

2. *Stemming*

Stemming merupakan tahapan preprocessing kata yang berfungsi untuk mengubah suatu term kedalam bentuk root [17]. Contoh dari *stemming* dapat dilihat pada Tabel 2.

TABEL 2. Proses *Stemming*

Data Input	Data Output
kuliah daring membawa berkah akhirnya dapat juga beasiswa	kuliah daring bawa berkah akhir dapat juga beasiswa

3. *Tokenizing*

Tahap *tokenization* adalah tahap dimana string input dipotong berdasarkan kata – kata yang membentuknya atau tahap dimana kalimat dibagi menjadi kata – kata [17]. Contoh dari *tokenizing* dapat dilihat pada Tabel 3.

TABEL 3. Proses *Tokenizing*

Data Input	Data Output
kuliah daring bawa berkah akhir dapat juga beasiswa	['kuliah', 'daring', 'bawa', 'berkah', 'akhir', 'dapat', 'juga', 'beasiswa']

D. Pembobotan TF – IDF

TF–IDF adalah metode yang mengintegrasikan term frequency (TF) dan inverse document frequency (IDF). Proses TF-IDF digunakan untuk mengevaluasi bobot relevansi term dari suatu dokumen terhadap keseluruhan dokumen dalam korpus. Frekuensi istilah ini dihitung menggunakan persamaan 1 dimana frekuensi term ke-i dan frekuensi kemunculan term ke-i dalam dokumen ke-j [20].

$$tf(i) = \frac{freq_i\,(d_j)}{\sum_{i=1}^{k} freq_i\,(d_j)} \qquad (1)$$

Sebaliknya, inverse document adalah logaritma dari rasio jumlah semua dokumen dalam korpus dengan jumlah dokumen yang secara matematis menggambarkan istilah yang dimaksud dalam persamaan 2 [20].

$$idf_1 = 1 + \log\left(\frac{N+1}{df(t)+1}\right) \qquad (2)$$

Alasan menambahkan '1' kedalam pembilang dan penyebut dan juga ke persamaan idf adalah untuk menjaga stabilitas numerik [21]. Nilai TF-IDF diperoleh dengan mengalikan keduanya yang dapat diformulasikan dengan persamaan 3 seperti dibawah ini [20].

$$(tf - idf)_{ij} = tf_i(d_j).idf_i \qquad (3)$$

Dimana d adalah dokumen ke-d, t adalah kata ke-t dari kata kunci, N adalah jumlah total dokumen, dan df adalah banyak dokumen yang mengandung kata-t

E. Complement Naïve Bayes

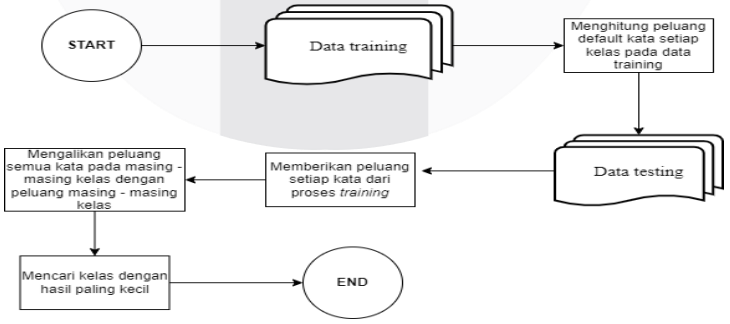
Algoritme Naïve Bayes adalah suatu metode pengklasifikasian berdasarkan probabilitas sederhana dan dirancang untuk digunakan dengan asumsi bahwa antar suatu kelas dengan kelas yang lain tidak saling bergantung atau independent [22]. Algoritme Complement Naive Bayes cocok digunakan untuk data yang tidak seimbang dan semua data latih dari kelas selain kelas c digunakan [23]. Perkiraan setiap kata dari kelas selain kelas c dapat dilihat pada persamaan 4 dibawah ini [23].

$$\hat{\theta}_{\tilde{c}i} = \frac{N_{\tilde{c}i} + \propto_i}{N_{\tilde{c}} + \propto} \qquad (4)$$

Untuk rumus dari Complement Naïve Bayes sendiri dapat dilihat pada persamaan 5 dibawah ini [23].

$$\mathrm{CNBC} = \mathrm{argmax}_c\left[\log \mathrm{P(c)} - \sum_i \mathrm{f_i} \log \hat{\theta}_{\tilde{c}i}\right] \qquad (5)$$

Dimana $N_{\tilde{c}i}$ adalah jumlah kemunculan kata i dalam suatu dokumen di kelas selain *c*, $N_{\tilde{c}}$ adalah jumlah total kemunculan kata di suatu kelas selain c, $\propto_i$ dan $\propto$ adalah parameter untuk *smoothing*, dan fi adalah banyak kata pada suatu kalimat. Cara kerja dari algoritme Complement Naïve Bayes dapat dilihat pada Gambar 1 dibawah ini.



GAMBAR 1. Proses Klasifikasi Menggunakan Algoritme Complement Naive Bayes

F. Pengujian Performansi Sistem

1. K – Fold Cross Validation

K – fold Cross Validation adalah suatu metode yang digunakan untuk mempartisi data menjadi dua yaitu data training dan data testing. Metode K-fold ini berfungsi untuk mengurangi bias ketika pengambilan sample data [24].

2. Confusion Matrix

Confusion matrix adalah suatu metode yang digunakan untuk menganalisis seberapa bagus classifier yang digunakan untuk mengenali tuple dari kelas yang berbeda [25]. Akurasi adalah metrik untuk model klasifikasi yang mengukur jumlah prediksi yang benar sebagai persentase dari jumlah total prediksi yang dibuat [26]. Untuk perhitungan akurasi dapat dilihat pada persamaan 6 dibawah ini [25].

$$\text{Akurasi} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \quad (6)$$

Pengukuran presisi dan recall juga sering digunakan dalam klasifikasi. Presisi berfungsi untuk melihat nilai sensitifitas atau nilai ketepatan suatu sistem antara informasi yang diberikan oleh sistem untuk menunjukkan secara benar suatu data kelas [30]. Rumus presisi dapat dilihat pada persamaan 7 dibawah ini [25].

$$\text{Presisi} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \quad (7)$$

Recall berfungsi untuk menunjukkan tingkat keberhasilan atau spesifisitas untuk mengetahui sebuah informasi secara benar tentang data suatu kelas [26]. Persamaan untuk *recall* dapat dilihat pada persamaan 8 dibawah ini [25].

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \quad (8)$$

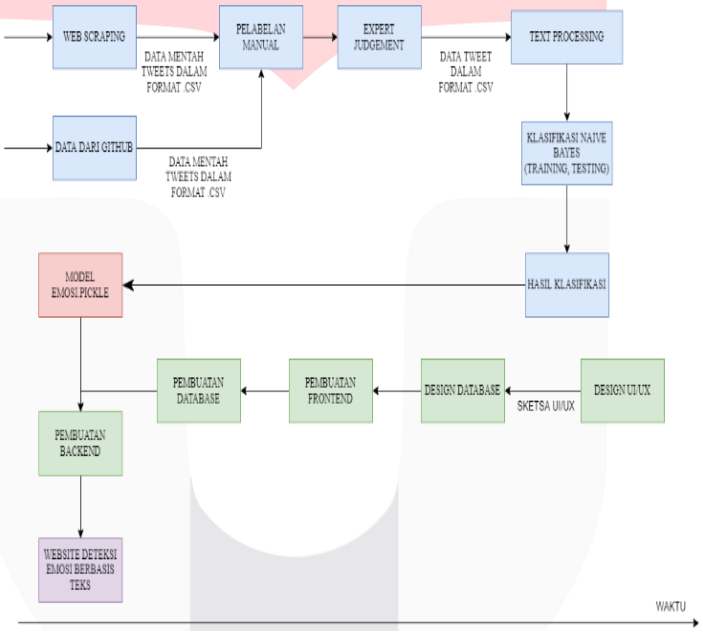
F1-*score* juga memiliki fungsi untuk melihat apakah model algoritme berjalan dengan baik [27]. Persamaan untuk f1-score dapat dilihat pada persamaan 2.9 dibawah ini [25].

$$\text{F1} - \text{score} = \frac{2 \text{ x presisi x } recall}{\text{presisi} + recall} \quad (9)$$

Dimana TP adalah True Positive, TN adalah True Negative, FP adalah False Positive, dan FN adalah False Negative.

III. METODE

1. Gambaran Umum Sistem



GAMBAR 2. Diagram Blok Sistem

Seperti yang dapat dilihat pada Gambar 2, proses yang dilakukan pada implementasi sistem ini terdiri dari tahap pengambilan data dari Twitter [28]. Pengambilan data dari Twitter akan dilakukan dengan web scraping menggunakan library Tweepy untuk mempermudah pengambilan data. Output dari pengambilan data tersebut akan disimpan dalam format .csv. Data yang duplikat akan dihapus dan kemudian diberikan pelabelan secara manual oleh ahli bahasa yang dikategorikan berdasarkan pelabelan emosi oleh Shaver, dkk [14]. Setelah itu, dilakukan text processing dan dilakukan pembobotan TF–IDF. Setelah itu, dilakukan klasifikasi menggunakan algoritme Complement Naïve Bayes. Setelah model sudah dibuat, model tersebut akan disimpan kedalam bentuk pickle dan diimplementasikan pada website yang dapat diakses oleh user.

2. Kebutuhan Data

Dataset dari Github terkait kuliah daring diambil dari rentang waktu data tersebut diambil pada 23 Agustus 2020 sampai 18 November 2020 [28]. Tidak hanya dari Github, data juga diambil dengan melakukan web scraping dari Twitted yang diambil dari rentang waktu November 2021 sampai Maret 2022. Total data yang digunakan dapat dilihat pada Tabel 4.

TABEL 4. Total Data yang Digunakan

Class	Happy	Angry	Love	Fear	Sadness	Total
Dataset Online Lecture						
3 Class of Emotion	1823	2047	1300	-	-	5170
4 Class of Emotion	1823	2047	1300	1651	-	6821
5 Class of Emotion	1671	1882	1215	1494	1738	8000

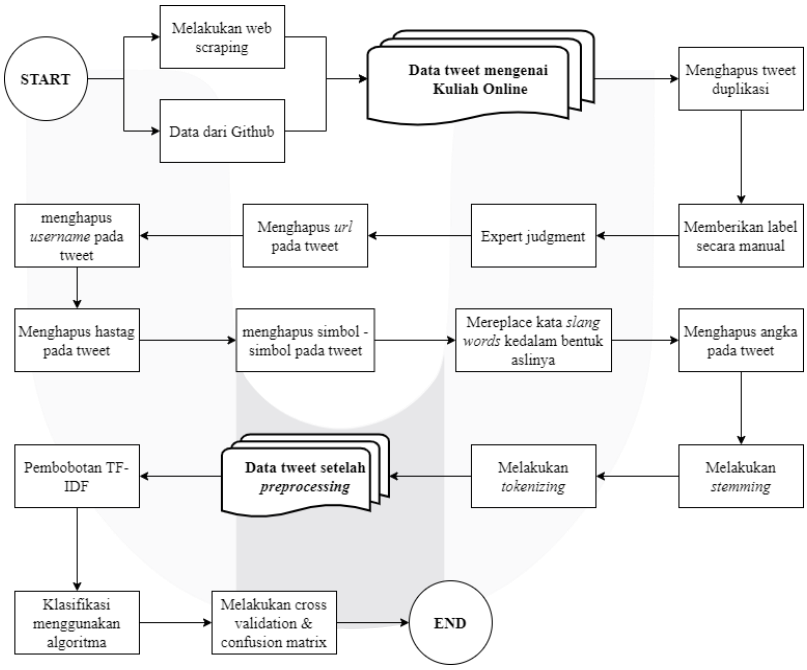
Contoh dari data yang digunakan pada Tugas Akhir ini dapat dilihat pada Tabel 5.

TABEL 5. Contoh Dataset yang Digunakan

Jenis	Tanggal	Tweets
Crawling Data	2022-03-17	@rentfess Ini jujur lucu banget tapi karena saya kuliah dan kerja masih daring dan WFH jadi gak bisa disambut pulang deh. ðŸ¤£ Tapi semoga ketemu yang cocok ya, Sender!
Github	2020-08-23	Eh bentar lgi september yakkk, yahhh bntr lgi mulai kuliahhh:(((males bngt kuliah online, bnyk tugasnyaa:((

3. Text Processing

Seperti pada Gambar 3, langkah pertama yang dilakukan pada text processing adalah menghapus URL, username, hastag, dan juga menghilangkan simbol pada *tweets*. Setelah itu, dilakukan case folding dan mengubah kata *slang words* menjadi kata baku. Setelah tahap tersebut selesai, dilakukan menghapus angka yang ada didalam tweets dan melakukan stemming. Terakhir, dilakukan tokenizing sebelum masuk tahap pembobotan TF-IDF.

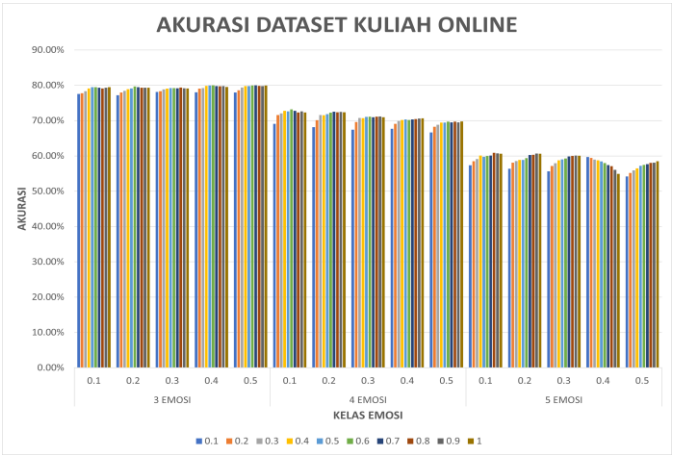


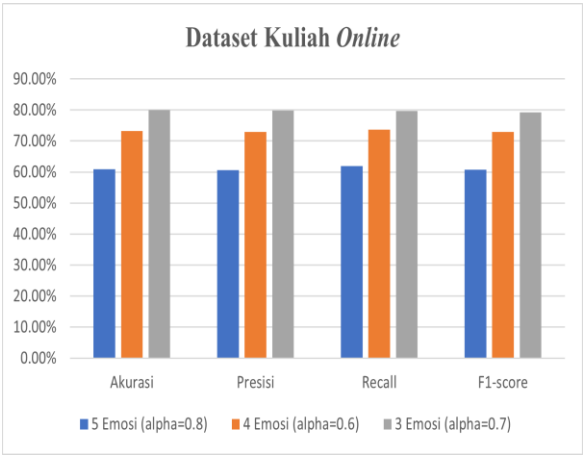
GAMBAR 3. Flowchart Text Processing

IV. HASIL DAN PEMBAHASAN

A. Pengujian Nilai Alpha dan Partisi Data pada Algoritme Complement Naïve Bayes

Untuk menguji performansi dari model algoritme yang sudah dibuat, dilakukan pencarian terhadap nilai alpha terbaik dari 0.1 sampai 1 dan juga mengubah partisi data untuk data uji dari 0.1 sampai 0.5 untuk mengetahui akurasi terbaik yang dihasilkan.





GAMBAR 4. Hasil Pengujian Dataset Kuliah *Online*

Seperti pada Gambar 4 dapat dilihat bahwa, hasil akurasi tertinggi berdasarkan nilai *alpha* dan juga pengujian partisi data untuk dataset kuliah *online* didapatkan pada kelas 3 emosi. Kelas tersebut mendapatkan akurasi sebesar 80% dengan nilai alpha 0.7 dan partisi dara untuk data testing 0.5.

B. Pengujian Performansi Sistem Algoritme Complement Naïve Bayes

Pengujian akan dilakukan dengan mencari nilai alpha dan partisi data terbaik. Pada pengujian K-Fold Cross Validation, pengujian dilakukan dengan nilai cv=10. Hasil dari pengujian K-Fold Cross Validation dapat dilihat pada Tabel 6 dan untuk Confusion Matrix dapat dilihat pada Tabel 7 dibawah ini.

TABEL 6. Hasil K-Fold Cross Validation

Emotion	CV (Hasil dalam %)										Mean	Accuracy
	1	2	3	4	5	6	7	8	9	10		
3 α=0.7 ts=0.5	77	77	78	79	82	77	82	80	82	82	79	80
4 α=0.6 ts=0.1	69	72	72	72	70	68	73	68	70	72	71	73.20
5 α=0.8 ts=0.1	64	61	64	61	64	55	53	54	55	52	58	60.87
Notes : TS = Test Size												

Berdasarkan tabel diatas, hasil dari mean dengan akurasi dari klasifikasi menggunakan algoritme Complement Naïve Bayes ketika dilakukan pengujian tidak jauh berbeda. Hal ini dapat dikatakan bahwa model yang digunakan dikatakan baik.

TABEL 7. Contoh Hasil Confusion Matrix 3 Emosi

Testing Num-	Training Data	Testing Data			Prediction Class		
					Happy	Angry	Love
1 (alpha = 0.7)	50%	50%	Actual Class	Happy	582	177	134
				Angry	63	945	30
				Love	65	48	541
Precision					81.97%	80.76%	76.73%
Recall					65.17%	91.04%	82.72%
Accuracy					:	80%	
Precision					:	79.82%	
Recall					:	79.64%	

Berdasarkan tabel diatas yang menunjukkan hasil dari pengujian menggunakan confusion matrix pada 3 kelas emosi, kita dapat menghitung akurasi, presisi, *recall*, dan f1-*score* untuk melihat seberapa baik model yang telah dibuat.

V. KESIMPULAN

A. Kesimpulan

Berdasarkan hasil penelitian dan pengujian yang telah dilakukan pada Tugas Akhir ini, maka dapat diambil kesimpulan bahwa:

1. Sistem deteksi emosi dalam bahasa Indonesia menggunakan algoritme Complement Naïve Bayes berbasis website berhasil mengklasifikasikan emosi pada teks yang berasal dari Twitter.
2. Performansi terbaik pada algoritme Complement Naïve Bayes pada dataset kuliah online yaitu pada 3 kelas emosi dengan perbandingan data *training* 50% dan data *testing* 50% mendapatkan akurasi 80%, presisi 79.82%, *recall* 79.64%, dan F1-*score* 79.27% dengan nilai alpha 0.7. Hasil akurasi yang dihasilkan pun tidak jauh berbeda saat diuji menggunakan *K-Fold Cross Validation* dimana mendapatkan rata – rata akurasi

berdasarkan cv=10 yaitu 79.9%. Hal ini membuktikan bahwa model yang digunakan pada Tugas Akhir dapat dikatakan baik. Selain itu, jumlah data dan jumlah kelas emosi yang digunakan juga akan mempengaruhi akurasi yang dihasilkan. Semakin sedikit kelas emosi yang digunakan, maka semakin baik hasil perfomasi sistem,

B. Saran

Berdasarkan hasil penelitian dan pengujian yang telah dilakukan pada Tugas Akhir ini, maka saran yang dapat diusulkan adalah:

1. Untuk meningkatkan akurasi, dapat dilakukan dengan mencoba menggunakan algoritme yang berbeda seperti menggunakan algoritme deep learning.
2. 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] R. N. Velarosdela, "Kilas Balik Kronologi Munculnya Kasus Pertama Covid-19 di Indonesia," 2 Maret 2021. [Online]. Available: https://megapolitan.kompas.com/read/2021/03/02/05300081/kilas-balik-kronologi-munculnya-kasus-pertama-covid-19-di-indonesia?page=all.

[2] V. B. Kusnandar, "Pengguna Internet Indonesia Peringkat ke-3 Terbanyak di Asia," 14 10 2021. [Online]. Available: https://databoks.katadata.co.id/datapublish/2021/10/14/pengguna-internet-indonesia-peringkat-ke-3-terbanyak-di-asia.

[3] Goleman, Daniel, H. A. Said, Becker dan Bieswanger, Kecerdasan Emosional Untuk Mencapai Puncak Prestasi, Jakarta: PT. Gramedia Pustaka Utama, 2001.

[4] H. Krishnan, M. S. Elayidom dan T. Santhanakrishnan, "Emotion Detection of Tweets using Naïve Bayes Classifier," *International Journal of Engineering Technology Science and Research*, pp. 457-462, 2017.

[5] P. T. Febrianto, S. Mas'udah dan L. A. Megasari, "Implementation Of Online Learning During The Covid-19 Pandemic On Madura Island, Indonesia," *International Journal of Learning, Teaching and Educational Research,* vol. 19, pp. 233-254, 2020.

[6] D. Glez-Pena, A. Lourenco, H. Lopez-Fernandez, M. Reboiro-Jato dan F. Fdez-Riverola, "Web scraping technologies in an API world," *Briefings In Bioinformatics,* pp. 1-10, 2013.

[7] R. Mitchell, Web Scraping with Python : Collecting Data from the Modern Web, United States of America: O'Reilly Media, 2015.

[8] "Twitter," 10 Januari 2013. [Online]. Available: https://www.techopedia.com/definition/4957/twitter. [Diakses 22 Maret 2022].

[9] C. Beaumont, "The Telegraph," 23 Februari 2010. [Online]. Available: https://www.telegraph.co.uk/technology/twitter/7297541/Twitter-users-send-50-million-tweets-per-day.html.

[10] A. Brown, "The Tricky Business Of Business Tweeting," 2 Maret 2011. [Online]. Available: https://www.irishtimes.com/life-and-style/people/the-tricky-business-of-business-tweeting-1.561138.

[11] "Developer Platform Twitter," [Online]. Available: https://developer.twitter.com/en/docs/twitter-api/getting-started/about-twitter-api#What's.

[12] P. Ekman, "Basic emotions. In T. Dalgleish & M. J. Power (Eds.)," *Handbook of Cognition and Emotion,* pp. 45-60, 1999.

[13] R. Plutchik dan H. Kellerman, A General Psycho Evolutionary Theory of Emotion, Amsterdam: Elsevier, 1980.

[14] P. R. S. a. U. Murdaya dan R. C. Fraley, "Structure Of The Indonesian Emotion Lexicon," *Asian Journal of Social Psychology,* pp. 201-224, 2001.

[15] C. D. Manning, P. Raghavan dan H. Schütze, An Introduction To Information Retrieval, Cambridge: Cambridge University Press, 2009.

[16] F. S. Jumeilah, "Penerapan Support Vector Machine (SVM) untuk Pengkategorian Penelitian," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi),* pp. 19-25, 2017.

[17] H. Benbrahim dan M. Bramer, "Text and Hypertext Categorization," *IFIP International Federation for Information Processing,* pp. 11-38, 2009.

[18] T. Septiana, "Slang Word Bahasa Inggris Ini Bisa Bikin Percakapan Makin Gaul dan Kekinian Lo," Lifestyle, 24 Agustus 2021. [Online]. Available: https://lifestyle.kontan.co.id/news/slang-word-bahasa-inggris-ini-bisa-bikin-percakapan-makin-gaul-dan-kekinian-lo?page=all#:~:text=Slang%20words%20merupakan%20bahasa%20gaul,percakapan%20atau%20tulisan%20non%2Dformal.. [Diakses 2022 Maret 23].

[19] N. A. Salsabila, Y. A. Winatmoko, A. A. Septiandri dan A. Jamal, "Colloquial Indonesian Lexicon," *International Conference on Asian*

Language Processing (IALP), pp. 226-229, 2018.

[20] N. K. Widyasanti, I. K. G. D. Putra dan N. K. D. Rusjayanthi, "Seleksi Fitur Bobot Kata dengan Metode TFIDF untuk Ringkasan Bahasa Indonesia," *Merpati*, pp. 119-126, 2018.

[21] V. Shah, "Hands-on implementation of TF-IDF from scratch in Python," Analytics India Magazine, 29 Desember 2020. [Online]. Available: https://analyticsindiamag.com/hands-on-implementation-of-tf-idf-from-scratch-in-python/#:~:text=TF%2DIDF%20is%20a%20method,IDF%20is%20Inverse%20document%20frequency. [Diakses 1 Juni 2022].

[22] M. N. Randhika, J. C. Young, A. Suryadibrata dan H. Mandala, "Implementasi Algoritma Complement dan Multinomial Naïve Bayes Classifier Pada Klasifikasi Kategori Berita Media Online," *Ultimatics : Jurnal Teknik Informatika*, pp. 19-25, 2021.

[23] J. D. M. Rennie, L. Shih, J. Teevan dan D. R. Karger, "Tackling the Poor Assumptions of Naive Bayes Text Classifiers," dalam *Proceedings of the Twentieth International Conference on Machine Learning (ICML-2003)*, Washington DC, 2003.

[24] E. Gokgoz dan A. Subasi, "Comparison of Decision Tree Algorithms for EMG Signal Classification Using DWT," *Biomedical Signal Processing and Control*, vol. 18, pp. 138-144, 2015.

[25] J. Han, M. Kamber dan J. Pei, Data Mining Concepts and Techniques, United States of America: Elsevier Inc., 2011.

[26] M. Azhari, Z. Situmorang dan R. Rosnelly, "Perbandingan Akurasi, Recall, dan Presisi Klasifikasi pada Algoritma C4.5, Random Forest, SVM dan Naive Bayes," *Jurnal Media Informatika Budidarma*, vol. 5, pp. 640-651, 2021.

[27] C. Goutte dan E. Gaussier, "A Probabilistic Interpretation of Precision, Recall and F-score, with Implication for Evaluation," *Proceedings of the European Colloquium*, pp. 345-359, 31 Agustus 2005.

[28] R. A. Saputra, "Sentiment Analysis Online Lectures in Indonesia," 10 Desember 2020. [Online]. Available: https://github.com/rifkyahmadsaputra/Sentiment-Analysis-Online-Lectures-in-Indonesia. [Diakses 10 Desember 2021].

[29] R. A. Saputra, "Github," 10 Desember 2020. [Online]. Available: https://github.com/rifkyahmadsaputra/Sentiment-Analysis-Online-Lectures-in-Indonesia/tree/main/data. [Diakses 12 Desember 2021].

[30] R. Feldman dan J. Sanger, The Text Mining Handbook Advanced Approaches In Analyzing Unstructured Data, United States of America: Cambridge University Press, 2007.