



Perbandingan Algoritma Naïve Bayes dan Support Vector Machine (SVM) dalam Analisis Sentimen Kebijakan Kemdikbudristek Tentang Kuota Internet Selama Pandemi Covid-19

Ulfa Khaira¹, Reni Aryani², Reza Wahyu Hardian³

^{1,2,3}Program Studi Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Jambi, Jl. Jambi - Muara Bulian No.KM. 15, Jambi, Indonesia.

ABSTRACT

Sentiment analysis is an activity that is used to analyze public opinion about an incident such as the Ministry of Education and Culture's internet assistance quota during the Covid-19 pandemic through one of the Twitter social media. Twitter is a microblogging platform that is used to write an opinion or opinion about an event that can be used as a source of data used. The Naïve Bayes method and Support Vector Machine (SVM) are methods with a Machine Learning approach that can be used to perform sentiment analysis on Kemdikbudristek policies regarding MoEC Quotas in the process of classifying a tweet based on its emotional level and knowing the accuracy comparison between the Naïve Bayes method and the Support Vector Machine (SVM). The results of the sentiment analysis process using the Naïve Bayes Algorithm and Support Vector Machine (SVM) based on public opinion, in this case Twitter users regarding the Ministry of Education and Culture Quota policies, resulted in a higher level of accuracy for the Support Vector Machine (SVM) than Naïve Bayes with an accuracy of 80%, for an average -the average precision value is 80.3%, recall is 80.3% and f1-score is 80.3%.

Keywords: Sentiment analysis, Naïve Bayes, Support Vector Machine (SVM)

ABSTRAK

Analisis sentimen adalah suatu bentuk kegiatan yang digunakan untuk menganalisis opini masyarakat tentang suatu kejadian seperti kuota bantuan internet Kemdikbudristek selama pandemi Covid-19 melalui salah satu media sosial Twitter. Twitter adalah suatu platform microblogging yang digunakan untuk menulis suatu opini atau pendapat tentang suatu peristiwa yang dapat dijadikan sebagai sumber data yang digunakan. Metode Naïve Bayes dan Support Vector Machine (SVM) merupakan metode dengan pendekatan Machine Learning yang dapat digunakan untuk melakukan analisis sentimen terhadap kebijakan Kemdikbudristek mengenai Kuota Kemdikbud dalam proses mengklasifikasi suatu tweet berdasarkan tingkat emosinya dan mengetahui perbandingan akurasi antara metode Naïve Bayes dengan Support Vector Machine (SVM). Hasil proses analisis sentimen menggunakan Algoritma Naïve Bayes dan Support Vector Machine (SVM) berdasarkan opini masyarakat dalam hal ini pengguna Twitter terhadap kebijakan Kemdikbudristek mengenai Kuota Kemdikbud menghasilkan tingkat akurasi Support Vector Machine (SVM) lebih tinggi dibandingkan Naïve Bayes dengan akurasi sebesar 80%, untuk rata-rata nilai presicion sebesar 80,3%, recall sebesar 80,3% dan f1-score sebesar 80,3%.

Kata Kunci: Analisis sentimen, Naïve Bayes, Support Vector Machine (SVM)

1. PENDAHULUAN

Covid-19 adalah jenis virus yang menyebabkan penyakit mulai dari gejala ringan sampai berat, penyakit jenis baru ini belum pernah diidentifikasi sebelumnya pada manusia. Dampak dari pandemi Covid-19 sudah semakin meluas hingga dirasakan oleh banyak kalangan di masyarakat, pandemi Covid-19 pastinya memberikan dampak diberbagai sektor kehidupan, termasuk sektor ekonomi, sosial, bisnis, transportasi, pariwisata dan pendidikan. Di Indonesia, dunia pendidikan mengalami perubahan yang cukup signifikan akibat pandemi, berbagai upaya telah dilakukan guna untuk mengurangi angka penyebaran Covid-19, salah satunya diterapkan sistem pendidikan berupa Pembelajaran Jarak Jauh (PJJ) yang masih sangat asing di Indonesia. Namun, dikarenakan sistem pembelajaran jarak jauh (PJJ) yang masih baru bagi warga Indonesia, tidak menutup kemungkinan menimbulkan beberapa masalah terkait pelaksanaannya, seperti keterbatasan internet yang dimiliki oleh masyarakat. Untuk mengatasi hal tersebut, Kementerian Pendidikan, Kebudayaan, Riset dan Teknologi (Kemdikbudristek) mengeluarkan dan menetapkan beberapa kebijakan baru untuk mengatasi permasalahan pendidikan dengan mempertimbangkan kondisi dimasa pandemi Covid-19 saat ini, salah satunya pemberian dana Bantuan Operasional Sekolah guna membiayai pembelajaran secara daring atau jarak jauh. Namun kebijakan tersebut tentunya menimbulkan berbagai opini pro atau kontra dari masyarakat terhadap kebijakan tersebut, seperti pemanfaatan media sosial twitter sebagai media masyarakat dalam memberikan tanggapan terhadap kebijakan kuota bantuan kemdikbud.

Pendapat masyarakat melalui media sosial twitter tentang kebijakan tersebut tentunya dapat dijadikan sebuah informasi yang berguna dan pengetahuan dalam pengambilan sebuah keputusan bagi pemerintah. Untuk mengetahui bagaimana tanggapan masyarakat perlu dilakukan analisis sentimen pada bidang ilmu text mining. Text mining adalah analisis teks dimana sumber data biasanya di peroleh dari suatu teks atau dokumen, dan tujuannya adalah mencari kata-kata yang dapat mewakili isi dari dokumen sehingga dapat dilakukan analisa kelas antar dokumen [1]. Analisis sentimen adalah suatu bidang text mining yang melakukan studi mengenai pendapat masyarakat, sentimen, evaluasi, perilaku, dan emosi terhadap suatu entitas [2]. Analisis sentiment merupakan suatu model klasifikasi yang dilakukan dengan berbagai bentuk permodelan, seperti pendekatan permodelan supervised learning. Supervised learning merupakan bagian dari machine learning yang membutuhkan data latih yang telah dilabelkan terlebih dahulu [3]. Oleh

karena itu pada penelitian ini, dilakukan analisis sentimen dengan pendekatan metode klasifikasi supervised learning menggunakan algoritma Naïve bayes dan Support Vector Machine (SVM).

Algoritma Naïve Bayes umumnya digunakan untuk melakukan prediksi sentimen yang muncul pada data yang masih belum memiliki sentimen, dan Naïve Bayes mampu mengelola data dalam jumlah besar dengan hasil akurasi yang tinggi [4]. Klasifikasi dengan SVM merupakan suatu usaha mencari hyperlane atau garis pembatas (decision boundary) terbaik yang berfungsi sebagai pemisah antara suatu kelas dengan kelas yang lainnya [5]. Terdapat beberapa penelitian tentang analisis sentiment menggunakan algoritma Naïve Bayes dan Support Vector Machine (SVM) antara lain penelitian analisis sentimen terkait lockdown pada media social twitter dengan menggunakan Support Vector Machine (SVM) dan Naïve Bayes [6]. Penelitian tersebut menghasilkan metode Support Vector Machine (SVM) memiliki nilai akurasi yang lebih tinggi sebesar 87% dibandingkan dengan menggunakan metode Naïve Bayes yang hanya mendapat nilai akurasi sebesar 81%. Ada juga penelitian perbandingan metode Naïve Bayes dan Support Vector Machine (SVM) pada analisis sentimen twitter. Pada penelitian ini penulis memaparkan bahwa metode Naïve Bayes merupakan metode yang baik digunakan dalam melakukan klasifikasi, hal itu dibuktikan dengan hasil akurasi yang didapatkan oleh penulis dalam melakukan penelitian sebesar 73,65% dibandingkan dengan SVM yang hanya mendapatkan nilai akurasi sebesar 70,20% dan menggunakan banyak data sebesar 2654 tweets [7].

Berdasarkan latar belakang diatas, peneliti akan menggunakan pendekatan metode Support Vector Machine (SVM) dan Naïve Bayes dalam melakukan klasifikasi sentimen data tweet, hal ini didorong untuk mengetahui perbandingan seberapa besar tingkat akurasi dalam melakukan klasifikasi data pada analisis sentiment dan mengetahui seberapa besar tanggapan masyarakat tentang kebijakan kuota internet.

2. TINJAUAN PUSTAKA

2.1. Analisis Sentimen

Analisis sentimen merupakan suatu bidang text mining yang melakukan studi mengenai pendapat masyarakat, sentimen, evaluasi, perilaku, dan emosi terhadap suatu entitas [8]. Tugas dasar analisis sentimen adalah mengelompokkan polaritas dari suatu teks yang terdapat dari suatu dokumen, kalimat atau suatu fitur entitas yang bersifat positif dan negatif [9]. Analisis sentiment dapat diklasifikasikan ke dalam kelas sentiment bersifat positif, negatif, dan netral. Sentimen Positif menurut Kamus Besar Bahasa Indonesia (Kamus Besar Bahasa Indonesia) sentimen positif merupakan reaksi atau sikap yang meningkatkan nilai seseorang atau sesuatu. Sentimen Negatif menurut Kamus Besar Bahasa Indonesia (Kamus Besar Bahasa Indonesia) sentimen positif merupakan reaksi atau sikap yang menurunkan nilai seseorang atau sesuatu. Sentimen netral menurut Kamus Besar Bahasa Indonesia (Kamus Besar Bahasa Indonesia) sentiment netral berarti tidak berpihak. Kalimat bersentimen betral merupakan opini yang berisi ungkapan yang tidak memiliki nilai yang bersifat positif maupun negatif.

2.2. Naïve Baiyes

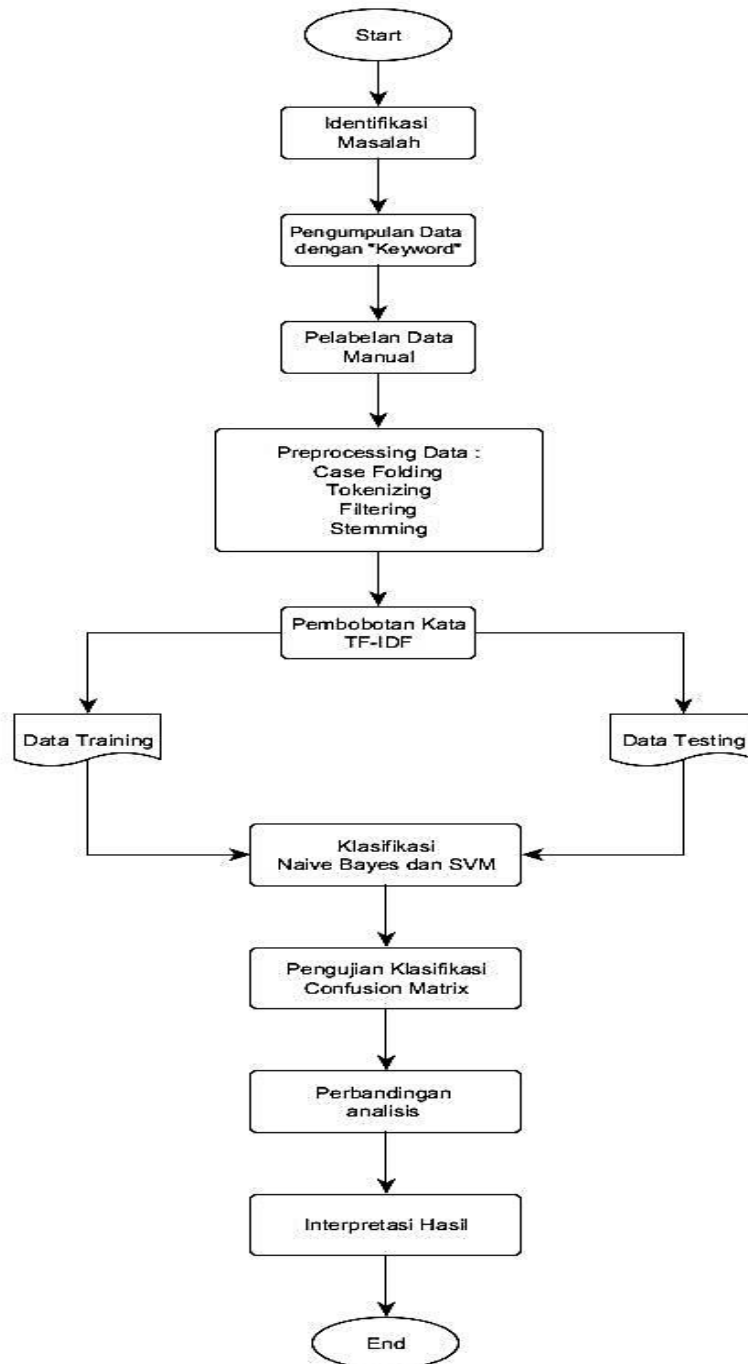
Naïve Baiyes Classifier merupakan pengklasifikasian dengan metode probabilitas dan statistik yang pertama kali ditemukan oleh ilmuwan inggris Thomas Bayes, yaitu memperkirakan peluang di masa depan berdasarkan pengalaman dimasa sebelumnya sehingga dikenal sebagai Teorema Bayes. Klasifikasi Naïve Bayes diprediksikan bahwa ada atau tidak ciri tertentu dari sebuah kelas tidak ada hubungannya dengan ciri dari kelas lainnya. Naïve Bayes berpotensi baik untuk mengklasifikasikan data karena kesederhanaannya [10].

2.3. Support Vector Machine (SVM)

Support Vector Machine (SVM) dikembangkan oleh Boser, Guyon dan Vapnik dan pertama kali di presentasikan pada tahun 1992 di Annual Workshop in Computational Learning Theory. Teknik ini termasuk salah satu jenis terpandu (supervised) karena memiliki target pembelajaran tertentu. Klasifikasi dengan SVM merupakan suatu usaha mencari hyperlane atau garis pembatas (decision boundary) terbaik yang berfungsi sebagai pemisah antara suatu kelas dengan kelas yang lainnya [11]. SVM melakukan pencarian nilai hyperlane yang paling maksimal dengan menggunakan Support Vector Machine dan nilai margin. SVM merupakan algoritma machine learning yang bekerja atas prinsip Struktural Risk Minimization (SRM) dengan tujuan menemukan hyperlane terbaik yang memisahkan dua buah class pada input space [12].

3. METODOLOGI PENELITIAN

Pada penelitian analisis sentiment ini dalam mencapai tujuan penelitian dilakukan dengan menggunakan support vector machine (svm) dan naïve bayes yang dilakukan dalam beberapa tahapan, berikut tahapan yang dilakukan dalam penelitian dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Analisis Sentimen

3.1. Identifikasi Masalah

Tahap Pertama dalam penelitian ini adalah melakukan identifikasi masalah. Identifikasi masalah yang didapatkan adalah bagaimana mengklasifikasikan opini pengguna twitter terhadap kebijakan kuota Kemdikbudristek ditengah pandemi Covid-19 dan mengetahui perbandingan tingkat akurasi dari penggunaan metode Support Vector Machine (SVM) dan Naïve Bayes pada analisis sentimen terhadap kebijakan kuota internet Kemdikbudristek.

3.2. Pengumpulan Data

Tahap selanjutnya adalah melakukan pengumpulan data opini twitter dengan menggunakan teknik scraping dengan bantuan dari website exportcomments.com.

3.3. Pelabelan Data

Setelah mendapatkan data dilakukan proses labeling atau pelabelan data dengan memberikan kategori kelas yang berbeda yaitu kelas positif, negatif dan netral.

3.4. Penanganan Imbalance Data

Setelah proses pelabelan data, selanjutnya dilakukan proses penanganan imbalance data atau proses penyeteraan jumlah data pada setiap kategori agar jumlah data pada setiap kategori berimbang. Random oversampling adalah salah satu metode yang dapat digunakan dalam menangani imbalance data [13]. Random Oversampling adalah metode yang digunakan untuk membantu mencapai distribusi kelas yang seimbang dengan replikasi sampel kelas minoritas [14].

3.5. Preprocessing Data

Setelah proses penanganan imbalance data selanjutnya data dilakukan tahapan preprocessing data yang terdiri dari proses Case Folding, Tokenizing, Filtering dan Stemming hingga data bersih.

3.6. Pembobotan Data

Setelah mendapatkan data bersih, selanjutnya dilakukan proses pembobotan data atau kata menggunakan metode TF-IDF (*Term Frequency-Inverse Document Frequency*). Proses pembobotan data digunakan untuk memberi nilai bobot *relevansi term* dari sebuah data terhadap keseluruhan dokumen yang ada [15]. Berikut rumus yang digunakan dalam menentukan bobot *Term Frequency-Inverse Document Frequency* (TF-IDF):

$$idf_t = \log \frac{td}{df} \quad (1)$$

$$W_{t,d} = tf_{t,d} \times idf_t \quad (2)$$

Keterangan :

$W_{t,d}$ = pembobotan TF-IDF

$tf_{t,d}$ = bobot kata t dalam setiap d

td = jumlah keseluruhan data yang ada

df = jumlah kemunculan kata pada semua data

idf_t = nilai IDF dari setiap kata yang akan dicari

3.7. Implementasi Klasifikasi Data

Setelah mendapatkan bobot dari setiap *term*, selanjutnya terlebih dahulu data dibagi menjadi 2 bagian yaitu 80% data latih dan 20% data uji. Data training digunakan untuk membuat proses pembuatan model klasifikasi dari metode SVM dan *Naïve Bayes*.

Adapun tahapan yang dilakukan pada pengklasifikasian dengan algoritma *Naïve Bayes* sebagai berikut :

1. *Dataset* hasil pembobotan TF-IDF
2. *Dataset* dibagi menjadi 80% data training dan 20% data latih dari keseluruhan data.
3. Baca data *training*.
4. Hitung jumlah dan probabilitas, jika data numerik maka dilakukan :
 - a. Cari nilai mean dan standar deviasi dari pada setiap parameter yang merupakan data numerik. Dengan persamaan:

$$\mu = \frac{\sum_{i=1}^n X_i}{n} \quad (3)$$

Atau

$$\mu = \frac{X_1 + X_2 + X_3 + \dots + X_n}{n} \quad (4)$$

Keterangan :

μ : Rata-rata hitung atau mean

X_i : Nilai Sampel ke -i

n : Jumlah sampel dan persamaan yang digunakan dalam menghitung nilai simpanan baku sebagai berikut:

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (X_i - \mu)^2}{n-1}} \quad (5)$$

Keterangan :

σ : Standar deviasi

X_i : Nilai Sampel ke -i

μ : Rata-rata hitung atau mean

n : jumlah sampel

b. cari nilai probabilitas dengan cara menghitung jumlah data yang sesuai dari kategori yang sama dibagi dengan jumlah data yang sesuai dari kategori.

5. Menghasilkan nilai dalam tabel *mean, standar deviasi*, dan probabilitas.
6. Solusi kemudian dihasilkan

Sedangkan pada algoritma *Support Vector Machine (SVM)* tahapan yang dilakukan pada pengklasifikasian sebagai berikut :

1. *Dataset* hasil pembobotan TF-IDF
2. Input data training yang telah di sediakan
3. Data Training
4. Menentukan inisiasi pada parameter – parameter dan melakukan perhitungan kernel RBF dengan persamaan :

$$K(x_i, x_j) = \exp \left(-\left| \frac{x_i - x_j}{2\sigma^2} \right|^2 \right) \quad (6)$$

5. Menghitung matriks Hessian dengan persamaan dibawah ini:

$$D_{ij} = y_i y_j (K(x_i, x_j) + \lambda^2) \quad (7)$$

Keterangan :

D_{ij} = elemen matriks data ke-ij

y_i = kelas atau label data ke-i

y_j = kelas atau label data ke-j

λ = turunan batas teoritis

1. Melakukan iterasi untuk iterasi 1,2,...,n dengan dimulai menghitung nilai E_i , $\delta\alpha_i$ dan α_i , dengan persamaan :

$$E_i = \sum_{j=1}^n \alpha_j D_{ij} \quad (8)$$

$$\delta\alpha_i = \min \{ \max[\gamma(1 - E_i), -\alpha_i], C - \alpha_i \} \quad (9)$$

$$\alpha_i = \alpha_i + \delta\alpha_i \quad (10)$$

Keterangan :

α_i = Nilai parameter alpha

D_{ij} = Hasil nilai matriks Hessian

C = nilai untuk parameter C

$\delta\alpha_i$ = Parameter delta alfa untuk ke-i

2. Mengulangi proses ke 2-3 hingga nilai $\delta\alpha_i$ hingga mencapai nilai maksimum maka iterasi dihentikan
3. Melakukan perhitungan mencari nilai bias (b) dengan menggunakan persamaan dibawah ini:

$$b = -\frac{1}{2} [w \cdot x^+ + w \cdot x^-] \quad (11)$$

4. Pengujian dengan data testing yang telah disediakan
5. Melakukan perhitungan keputusan, dengan persamaan sebagai berikut:

$$h(x) = \begin{cases} +1, & \text{if } w \cdot x + b \geq 0 \\ -1, & \text{if } w \cdot x + b \leq 0 \end{cases} \quad (12)$$

$$\text{sign } h(x) = w \cdot x + b \quad (13)$$

$$\text{sign } h(x) = \sum_{i=1}^m \alpha_i y_i K(x_i, x_j) + b \quad (14)$$

Model klasifikasi disimpan dan data uji digunakan untuk menguji model klasifikasi yang telah dibuat sehingga didapatkan hasil dari sentiment analisis untuk mengetahui seberapa besar tanggapan masyarakat terhadap kebijakan Kuota Kemdikbudristek pada masa pandemi Covid-19.

3.8. Pengujian Klasifikasi

Setelah proses klasifikasi data selesai selanjutnya dilakukan pengujian model klasifikasi dengan menggunakan *Confusion matrix* yang bertujuan untuk membandingkan hasil klasifikasi yang dilakukan oleh sistem dengan hasil klasifikasi yang seharusnya. Tahap terakhir ini adalah menghitung nilai *akurasi*, *Recall*, *Precision* dan *f1-Score* dari model klasifikasi sehingga dapat diketahui perbandingan tingkat akurasi pada model klasifikasi. Dengan persamaan perhitungan akurasi, *precision*, *recall* dan *f1-score* sebagai berikut.

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \quad (15)$$

$$Precision = \frac{TP}{TP+FP} \quad (16)$$

$$recall = \frac{TP}{TP+FN} \quad (17)$$

$$f1 - Score = 2 \times \frac{(precision \times Recall)}{(precision + Recall)} \quad (18)$$

4. HASIL DAN PEMBAHASAN

4.1. Pengumpulan Data

Kegiatan pengumpulan data tweet berbahasa Indonesia dengan pencarian kata kunci “kuota kemdikbud” dan “kuota belajar kemdikbud” yang dilakukan dengan teknik pengumpulan data yaitu scraping data twitter melalui website exportcomment.com. Pertama-tama dilakukan terlebih dahulu pencarian data sesuai dengan kata kunci yang diinginkan pada menu pencarian pada aplikasi twitter lalu dilakukan pengambilan link dari pencarian data sesuai kata kunci, setelah itu link dimasukkan kedalam website exportcomment.com dan sistem akan melakukan scraping data secara otomatis sesuai dengan link yang dimasukkan. Dari proses scraping data twitter diperoleh data sebanyak 1078 data tanpa batasan hasil scraping.

4.2. Pelabelan Data

Pelabelan secara manual oleh Kantor Bahasa Provinsi Jambi Kementerian Pendidikan dan Kebudayaan – Riset dan Teknologi. Dari proses tersebut, didapatkan hasil pelabelan dengan nilai sentiment pada pelabelan manual sebesar 265 data bernilai positif, 287 data bernilai negatif dan 525 data bernilai netral. Tabel 1 menyajikan contoh pelabelan kelas positif, negatif dan netral.

Tabel 1. Contoh Pelabelan

Kelas	Tweet
Positif	@anican yaAllah bersyukur bgt dapet kuota kemdikbud 10gb sama 90gb tp ini gmn ngabisinnya sampe february 😊
Negatif	@Kemdikbud_RI @DPR_RI Pak menteri terima kasih atas kuota belajar nya tapi mohon maaf sebelumnya untuk bantuan kuota belajar. Kecepatan untuk mengakses internet sangat susah sekali. Untuk browsing dan elearning juga susah pak 🙏
Netral	@CALLMETEMPO12 Halo, Kakak Tempo. Mimin bantu info ya, Kuota Internet Kemdikbud bulan keempat akan terbit pada tanggal 15

4.3. Penanganan Imbalance Data

Setelah melakukan proses pelabelan data dilakukan proses penanganan imbalance data yang dimana dari hasil pelabelan data didapatkan ketidakseimbangan class antar data, oleh karena itu pada penelitian ini perlu dilakukan proses penanganan ketidakseimbangan class antar data dengan menggunakan metode Random Oversampling. Dari proses penanganan imbalance data didapatkan jumlah data dengan label netral sebesar 559 data, positif 520 data dan negatif 511 data.

4.4. Preprocessing Data

Case folding adalah proses menyeragamkan huruf pada suatu kalimat dari huruf kapital menjadi huruf standar atau kecil semua [16]. Berikut merupakan contoh data yang dilakukan proses *case folding*.

Tabel 2. Hasil Case Folding

Tweet	Case Folding
@Itjen_Kemdikbud Kuota bantuan Kemendikbud kenapa kok dikurangi pak?@Itjen_Kemdikbud padahal butuh bgt itung"	@itjen_kemdikbud kuota bantuan kemendikbud kenapa kok dikurangi pak?@itjen_kemdikbud padahal butuh bgt itung"

Tokenizing dalam penelitian ini memecahkan kata per kata pada sebuah kalimat atau dokumen yang telah melewati tahap case folding berdasarkan tiap kata yang menyusunnya dan menghilangkan URL, @mention dan hashtag [17].

Tabel 3. Hasil Tahap Tokenizing

Tweet	Tokenizing
@itjen_kemdikbud kuota bantuan kemendikbud kenapa kok dikurangi pak?	['kemdikbud', 'kuota', 'bantuan', 'kemendikbud', 'kenapa', 'kok', 'dikurangi', 'pak']

Proses filtering atau disebut dengan *Stopword Removal* yang bertujuan untuk membuang kata yang tidak memiliki makna dengan bantuan library *stopword* bahasa Indonesia yang telah disediakan python3 (Kaur & Buttar, 2018).

Tabel 4. Hasil Filtering

Tweet	Filtering
['kuota', 'kemendikbud', 'kok', 'sekarang', 'cuma', 'gb', 'sih', 'dapatnya']	['kuota', 'kemendikbud', 'dapatnya']

Stemming data adalah suatu proses yang terdapat dalam sistem IR yang mentransformasi kata-kata yang terdapat dalam suatu dokumen ke kata-kata akarnya (root word) [18]. Tahap stemming dilakukan dengan menggunakan bantuan library *Sastrawi* yang terdapat pada bahasa pemrograman Python3.

Pembobotan dilakukan dengan menggunakan *library Skicit Learn* yang dapat digunakan dalam *Machine Learning*. Pada *library* tersebut terdapat algoritma *TF-IDF Vectorizer* yang dapat mengubah fitur teks menjadi sebuah representasi *vector*. Hasil dari *vectorisasi* diperoleh 1590 data mempunyai 1167 kata. Sebelum dilakukan klasifikasi, data dibagi menjadi data training dan testing dengan menggunakan *library train_test_split* pada *library scikit-learn* yang telah disediakan oleh python3 dengan pembagian data 80:20 dari jumlah data. Dari proses tersebut didapatkan jumlah data training sebesar 1272 data dan data testing 318 data.

4.5. Implementasi Algoritma Klasifikasi

Selanjutnya dilakukan proses pembuatan model dengan data training dengan menggunakan menggunakan algoritma Naïve Bayes dan Support Vector Machine (SVM). Penerapan metode Algoritma Naïve Bayes dilakukan dengan memanfaatkan *library Multinomial*

precision sebesar 80,3%, recall sebesar 80,3% dan f1-Score sebesar 80,3%. Dari hasil analisis tersebut didapatkan nilai akurasi, precision, recall dan f1-Score pada metode SVM lebih tinggi dibandingkan dengan metode Naïve Bayes.

DAFTAR PUSTAKA

- [1] R. Talib, M. Kashif, S. Ayesha, and F. Fatima, "Text Mining: Techniques, Applications and Issues," *Int. J. Adv. Comput. Sci. Appl.*, vol. 7, no. 11, pp. 414–419, 2016, doi: 10.14569/IJACSA.2016.071153.
- [2] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," *Ain Shams Eng. J.*, 2014, doi: 10.1016/j.asej.2014.04.011.
- [3] S. VijayGaikwad, A. Chaugule, and P. Patil, "Text Mining Methods and Techniques," *Int. J. Comput. Appl.*, vol. 85, no. 17, pp. 42–45, Jan. 2014, doi: 10.5120/14937-3507.
- [4] S. Xu, "Bayesian Naïve Bayes classifiers to text classification," *J. Inf. Sci.*, 2018, doi: 10.1177/0165551516677946.
- [5] S. Sulova, "USING TEXT MINING TO CLASSIFY RESEARCH PAPERS," in *International Multidisciplinary Scientific GeoConference Surveying Geology and Mining Ecology Management, SGEM*, Jun. 2017, vol. 17, no. 21, pp. 647–654. doi: 10.5593/sgem2017/21/S07.083.
- [6] M. D. Alizah, A. Nugroho, U. Radiyah, and W. Gata, "Sentimen Analisis Terkait 'Lockdown' pada Sosial Media Twitter," *CSRID (Computer Sci. Res. Its Dev. Journal)*, 2021, doi: 10.22303/csrtd.12.3.2020.143-149.
- [7] M. I. Fikri, T. S. Sabrila, and Y. Azhar, "Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter," *SMATIKA J.*, vol. 10, no. 02, pp. 71–76, Dec. 2020, doi: 10.32664/smatika.v10i02.455.
- [8] B. Liu, "Sentiment analysis and opinion mining," *Synth. Lect. Hum. Lang. Technol.*, 2012, doi: 10.2200/S00416ED1V01Y201204HLT016.
- [9] M. Kaya, G. Fidan, and I. H. Toroslu, "Sentiment analysis of Turkish political news," in *Proceedings - 2012 IEEE/WIC/ACM International Conference on Web Intelligence, WI 2012*, 2012. doi: 10.1109/WI-IAT.2012.115.
- [10] S. L. Ting, W. H. Ip, and A. H. C. Tsang, "Is Naïve bayes a good classifier for document classification?," *Int. J. Softw. Eng. its Appl.*, vol. 5, no. 3, pp. 37–46, 2011.
- [11] K. T. Wahyuni, I. M. O. Widyantara, and N. D. Wirastuti, "Deteksi Tipe Modulasi Digital Pada Automatic Modulation Recognition Menggunakan Support Vector Machine dan Conjugate Gradient Polak Ribiere-Backpropagation," *Maj. Ilm. Teknol. Elektro*, vol. 18, no. 2, Aug. 2019, doi: 10.24843/MITE.2019.v18i02.P18.
- [12] F. Rahutomo, I. F. Rozi, and H. Setiyono, "Implementasi Support Vector Machine pada Analisa Sentimen Twitter Berdasarkan Waktu," *J. TAM (Technology Accept. Model.*, vol. 10, pp. 83–88, 2019.
- [13] R. Mohammed, J. Rawashdeh, and M. Abdullah, "Machine Learning with Oversampling and Undersampling Techniques: Overview Study and Experimental Results," in *2020 11th International Conference on Information and Communication Systems (ICICS)*, Apr. 2020, pp. 243–248. doi: 10.1109/ICICS49469.2020.239556.
- [14] S. M. Abd Elrahman and A. Abraham, "A Review of Class Imbalance Problem," *J. Netw. Innov. Comput.*, vol. 1, pp. 332–340, 2013, [Online]. Available: www.mirlabs.net/jnic/index.html
- [15] S. Qaiser and R. Ali, "Text Mining: Use of TF-IDF to Examine the Relevance of Words to Documents," *Int. J. Comput. Appl.*, 2018, doi: 10.5120/ijca2018917395.
- [16] B. R. Aditya, "Penggunaan Web Crawler Untuk Menghimpun Tweets dengan Metode Pre-Processing Text Mining," *J. INFOTEL - Inform. Telekomun. Elektron.*, vol. 7, no. 2, p. 93, Nov. 2015, doi: 10.20895/infotel.v7i2.35.
- [17] M. Straka and J. Straková, "Tokenizing, POS tagging, lemmatizing and parsing UD 2.0 with UDPipe," in *CoNLL 2017 - SIGNLL Conference on Computational Natural Language Learning, Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, 2017. doi: 10.18653/v1/k17-3009.
- [18] D. S. Maylawati, W. B. Zulfikar, C. Slamet, M. A. Ramdhani, and Y. A. Gerhana, "An Improved of Stemming Algorithm for Mining Indonesian Text with Slang on Social Media," in *2018 6th International Conference on Cyber and IT Service Management (CITSM)*, Aug. 2018, pp. 1–6. doi: 10.1109/CITSM.2018.8674054.
- [19] S. Y. Ying, P. Keikhosrokiani, and M. P. Asl, "Opinion Mining on Viet Thanh Nguyen's The Sympathizer Using Topic Modelling and Sentiment Analysis," *J. Inf. Technol. Manag.*, vol. 14, pp. 163–183, 2022, doi: 10.22059/jitm.2022.84895.