ANALISIS SENTIMEN NETIZEN TERHADAP ISU PEMBABATAN HUTAN ADAT PAPUA MELALUI TAGAR #ALLEYESONPAPUA MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE

Agustinus Manwombreidy Kafiar, Supatman
Program Studi Informatika, Fakultas Teknologi Informasi
Universitas Mercu Buana Yogyakarta
Jl. Jembatan Merah, No. 84C, Gejayan, Yogyakarta, Indonesia
aguz31.kafiar@gmail.com

ABSTRAK

Isu pembabatan hutan adat di Papua untuk perkebunan sawit menjadi topik sedang ramai diperbincangkan di media sosial, sehingga memahami opini netizen yang pro dan kontra terhadap isu ini menjadi sangat penting terutama di medial sosial X. Penelitian ini bertujuan menganalisis dan klasifikasi sentimen netizen terhadap isu pembabatan hutan adat di Papua melalui kata kunci tagar #AllEyesOnPapua di media sosial X. Penelitian ini menggunakan algoritma *Support Vector Machine (SVM)* dengan menggunakan bahasa pemrograman *Phyton* di website *Google Colaboratory* dengan menggunakan *API key Twitter*. Dari data yang dikumpulkan menggunakan *crawling data* antara tanggal 25 Mei sampai dengan 20 Juni 2024, terdapat 1270 tweet dan tersisa 1149 tweet setelah melalui tahap *preprocessing* hingga klasifikasi sentimen. Dengan menggunakan algoritma *Support Vector Machine* (SVM) data opini netizen dapat diklasifikasikan menjadi tigas kelas sentimen yaitu negatif, netral, dan positif. Hasil evaluasi berupa *Confusion Matrix* yang menunjukkan nilai *accuracy* sebesar 67%, dengan *precision* untuk sentimen positif sebesar 73%, *recall* 64%, dan *f1-score* 68%. Temuan ini menunjukkan bahwa model SVM cukup efektif dalam mengidentifikasi sentimen positif dibandingkan dengan sentimen negatif dan netral.

***Kata kunci**: #AllEyesOnPapua, Analisis Sentimen, Support Vector Machine*

1. PENDAHULUAN

Tanah Papua adalah rumah bagi lebih dari 250 kelompok etnis, dengan semua lahan diakui sebagai milik masyarakat adat. Seperti wilayah Melanesia lainnya, kelompok etnis di Papua terorganisir dalam marga yang memiliki hubungan spiritual mendalam dengan tanah, tanaman, dan hewan. Lahan marga bukan hanya bagian dari identitas mereka, tetapi juga sumber makanan dan kebutuhan dasar lainnya [1].

Namun, pada akhir Mei hingga awal Juni 2024, isu pembabatan hutan adat Boven Digoel di Papua viral di media sosial X dan Instagram dengan tagar #AllEyesOnPapua. Hingga 4 Juni 2024 tagar ini diposting sebanyak 2,8 juta kali di Insta Story. Tagar ini menunjukkan solidaritas netizen Indonesia terhadap gugatan hukum masyarakat adat Awyu dan Moi di Papua yang diwakili oleh Hendrikus Woro. Menurut masyarakat, pemerintah dan perusahaan sawit, seperti PT Indo Asiana Lestari (IAL) dan PT Sorong Agro Sawitindo, dianggap menghancurkan hutan adat Boven Digoel. Pemerintah memberikan izin penggunaan lahan adat suku Awyu dan suku Moi seluas 36.094 hektar untuk perkebunan sawit tanpa memikirkan nasib masyrakat adat.

Pada gugutan di PTUN Provinsi Papua untuk menuntut pencabutan izin yang dianggap bertentangan dengan kearifan lokal, Majelis Hakim PTUN menolak gugatan tersebut dan menyatakan izin sudah sesuai prosedur Amdal, meski pemerintah Provinsi Papua tidak transparan dan tidak melibatkan masyarakat sehingga sesuai PP No. 22/2021. Hendrikus kemudian pergi ke Jakarta untuk mengajukan kasasi di Mahkamah Agung. Selain jalur hukum, perwakilan suku Awyu dan Moi melakukan demonstrasi, doa bersama, dan ritual adat di depan Gedung MA, yang akhirnya diposting di media sosial dengan tagar #AllEyesOnPapua.

Viralnya tagar #AllEyesOnPapua memiliki dampak signifikan terhadap opini publik di media sosial X mengenai kasus ini. Setelah melakukan beberapa observasi di media sosial X, peneliti menemukan adanya isu negatif dan positif yang dapat mempengaruhi masyarakat. Oleh karena itu, tujuan dari penelitian ini adalah untuk menganalisa sentimen netizen berdasarkan opini-opini yang mereka sampaikan dalam media sosial X.

Analisis sentimen atau *opinion mining* adalah proses otomatis untuk memahami dan mengekstrak sentimen dari data teks, dengan tujuan menentukan apakah opini tersebut bersifat negatif atau positif [2]. Analisis sentimen mengumpulkan data dari berbagai platform internet untuk memahami emosi dalam teks ulasan. Tujuannya adalah memprediksi dan mengevaluasi suasana publik serta perasaan netizen secara otomatis, memberikan gambaran yang jelas tentang opini publik untuk pengambilan keputusan berbasis data [3].

Metodologi yang dipergunakan pada penelitian ini melingkupi tahap *crawling* data, *preprocessing*, dan klasifikasi sentimen. *Crawling* data diambil menggunakan pencarian pendapat netizen pada media sosial X. Sesudah terkumpul, selanjutnya melakukan proses mengolah data seperti preprocessing dan ekstrasi fitur untuk mempersiapkan data sebelum

dilakukan klasifikasi sentimen. Selanjutnya, yaitu sentimen analysis menggunakan algoritma *Support Vector Machine* (SVM) untuk mengklasifikasi data opini netizen menjadi negatif, netral, dan positif. Hasil analisis sentimen akan dianalisis dan diinterpretasikan untuk mendapat pemahaman yang baik dari opini dan pandangan masyarakat terhadap isu pembabatan Hutan Adat Boven Digoel, Papua melalui tagar #AllEyesOnPapua menggunakan algoritma *Support Vector Machine* SVM.

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Setiap penelitian memiliki kekhasan tersendiri meskipun menggunakan teori yang sama. Penelitian ini berfokus pada Analisis Sentimen Netizen Terhadap Isu Pembabatan Hutan Adat Boven Digoel Papua melalui tagar #AllEyesOnPapua di Media Sosial X Menggunakan Algoritma Support Vector Machine (SVM). Keterkaitannya dengan penelitian sebelumnya menunjukkan bagaimana metode SVM dapat diterapkan dalam berbagai konteks untuk menganalisis sentimen publik terhadap isu-isu yang berbeda.

Dalam jurnal yang berjudul Analisis Sentimen *Hate Speech* Pada Portal Berita Online Menggunakan *Support Vector Machine* (SVM) pada tahun 2020 mendeteksi ujaran kebencian di portal berita online Indonesia dengan akurasi 53,88% [4]. Selanjutnya, dalam penelitian yang berjudul "Analisis Sentimen Pada Maskapai Penerbangan di Platform Twitter Menggunakan Algoritma Support Vector Machine (SVM)" mengkaji opini pengguna Twitter tentang maskapai penerbangan dengan akurasi 84,37% [5]. Jurnal lain yang berjudul Analisis Sentimen Twitter Menggunakan *Metode Support Vector Machine* (SVM) pada Kasus Benih Lobster 2020 mengklasifikasikan sentimen terhadap kasus korupsi benih lobster dengan akurasi 84,21% [6]. Sedangkan dalam penelitian yang berjudul Analisis Sentimen Transportasi Online Menggunakan *Support Vector Machine* Berbasis *Particle Swarm Optimization* pada tahun 2020 menunjukkan akurasi sebesar 96,04% [7].

Penelitian ini diharapkan memberikan kontribusi signifikan dalam memahami respons netizen terhadap isu lingkungan dan sosial di media sosial, serta memperkuat efektivitas SVM dalam analisis sentimen.

2.2. Analisis Sentimen

Analisis sentimen, atau sering disebut *opinion mining*, adalah proses otomatis untuk memahami, mengekstrak, dan mengolah data teks guna mengidentifikasi sentimen dalam sebuah kalimat opini[8].

Menurut Turney, *opinion mining* melibatkan proses otomatis untuk memahami, mengekstrak, dan mengolah data teks guna mendapatkan informasi sentimen dari sebuah kalimat opini[9].

Tujuan analisis sentimen adalah untuk mengevaluasi pendapat atau kecenderungan opini seseorang mengenai suatu topik atau objek, apakah opini tersebut bersifat negatif atau positif.

Berita online biasanya berupa data teks digital yang tidak terstruktur [10]. Oleh karena itu, data tersebut perlu diklasifikasikan untuk mendapatkan struktur yang lebih jelas. Salah satu cara adalah dengan menggunakan teknik *opinion minin*g, yang memetakan sentimen dalam teks untuk memudahkan analisis. Dengan *opinion mining*, kita dapat mengetahui apakah opini bersifat positif atau negatif.

Opinion mining bisa dilihat sebagai gabungan antara text mining dan pemrosesan bahasa alami *(Natural Language Processing)* [11].

Text mining sebagai proses ekstraksi data dari teks, yang sumber datanya biasanya berasal dari dokumen[12]. Sementara NLP adalah cabang dari kecerdasan buatan yang berfokus pada pengkajian hubungan kompleks antara komputer dan bahasa manusia[13].

2.3. *Text Mining*

Text mining adalah proses mengekstrak informasi dari teks yang tidak terstruktur menjadi teks yang bermanfaat dan terstruktur. Tujuan utamanya adalah untuk menemukan pola, tren, atau wawasan yang tersembunyi dalam teks, yang bisa dipergunakan dalam pengambilan keputusan yang lebih baik atau analisis lebih lanjut. Teknik-teknik text mining sering melibatkan penggunaan pemrosesan bahasa alami (NLP), analisis statistik, dan pembelajaran mesin[14]. Text mining digunakan dalam berbagai aplikasi seperti analisis sentimen, pengelompokan dokumen, klasifikasi teks, dan ekstraksi informasi.

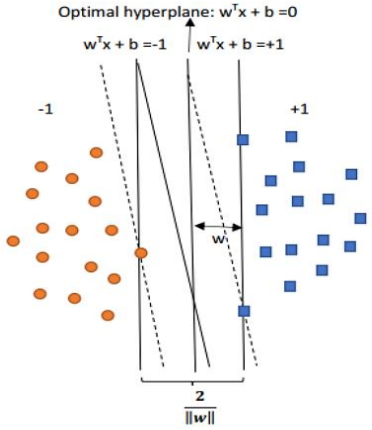
2.4. *Support Vector Machine (SVM)*

Salah satu metode dalam menganalisis sentimen netizen adalah menggunakan metode *Support Vector Machine*. SVM pertama kali diperkenalkan oleh Vapnik, Boser, dan Guyon pada tahun 1992. Meskipun SVM adalah teknik yang relatif baru dibandingkan dengan teknik lainnya, SVM telah menunjukkan performa yang unggul di berbagai bidang aplikasi seperti bioinformatika, pengenalan tulisan tangan, klasifikasi teks, dan klasifikasi diagnosis penyakit [15]. Menurut Maimon, Support Vector Machines (SVM) secara sederhana didefinisikan sebagai kumpulan metode terkait untuk pembelajaran mesin, yang dapat digunakan untuk masalah klasifikasi dan regresi. Dengan demikian, SVM dapat diterapkan untuk analisis sentimen netizen [16].

SVM adalah algoritma yang bekerja dengan menggunakan pemetaan nonlinier untuk mengubah data pelatihan asli ke dalam dimensi yang lebih tinggi. Di dimensi baru ini, algoritma akan mencari hyperplane yang dapat memisahkan data secara linier. Dengan pemetaan nonlinier yang tepat ke dimensi lebih tinggi, data dari dua kelas dapat selalu dipisahkan oleh hyperplane tersebut.

Dalam konteks penelitian ini adalah mencari du akelas pemisah antara netizen yang mendukung atau menolak atau bahkan mungkin netral. SVM mencapai ini dengan menggunakan support vector dan margin [17].

Tujuannya adalah menemukan fungsi pemisah (hyperplane) terbaik di antara banyak kemungkinan fungsi untuk memisahkan dua jenis objek. Hyperplane terbaik adalah yang berada di tengah-tengah antara dua kelompok objek dari dua kelas [18].



Gambar 1. Hyperplane yang terletak di tengah-tengah antara dua set obyek dari dua kelas

Kemudian merumuskan problem optimasi SVM untuk klasifikasi linear sebagai berikut [19]:

$$\min \frac{1}{2} \|\omega\|^2$$

Subject to

$$y_i(wx_i+b) \geq 1, i=1,\ldots, \lambda$$

Keterangan:
Xi: data input
Yi: keluaran dari data xi
W dan b: parameter-parameter q yang dicari nilainya.

Dalam rumus di atas, tujuan yang ingin dicapai adalah meminimalkan fungsi objektif 1/2 $\|w\|^2$ atau memaksimalkan kuantitas $\|w\|^2$ atau w^Tw dengan mempertimbangkan batasan $y_1(wx_i+b) \geq 1$. Jika output data $y_1 = +1$, maka batasannya menjadi $(wx_i + b) \geq 1$. Sebaliknya, jika $y_i=-1$, batasannya berubah menjadi $(wx_i+b) \leq -1$. Dalam kasus yang tidak feasible, di mana beberapa data mungkin tidak dapat dikelompokkan dengan benar, rumusan batasan ini akan disesuaikan menjadi:

$$\min \frac{1}{2} \|\omega\|^2 + C\sum_{i=1}^{\lambda} t_i$$

Subject to

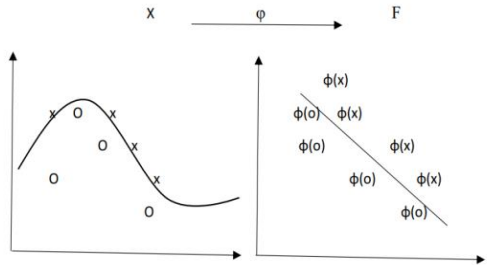
$$y_i(wx_i+b) + t_i \geq 1$$

$$t_i \geq 0, i=1,\ldots, \lambda$$

Keterangan:
ti: variabel slack.

Dengan formulasi ini, kita ingin memaksimalkan margin antara dua kelas dengan meminimalkan $\|w\|^2$. Dalam pendekatan ini, kita berusaha meminimalkan kesalahan klasifikasi (misclassification error) yang dinyatakan dengan variabel slack t_i, sambil tetap memaksimalkan margin $\|w\|^2$ Variabel slack t_i digunakan untuk mengatasi kasus ketidaklayakan (infeasibility) dari batasan $y_i(wx_i + b) \geq 1$ dengan cara memberikan penalti pada data yang tidak memenuhi batasan tersebut. Untuk meminimalkan nilai t_i ini, penalti diberikan dengan menerapkan konstanta ongkos C. Vektor w tegak lurus terhadap fungsi pemisah: wx + b = 0. Konstanta b menentukan lokasi fungsi pemisah relatif terhadap titik asal (origin).

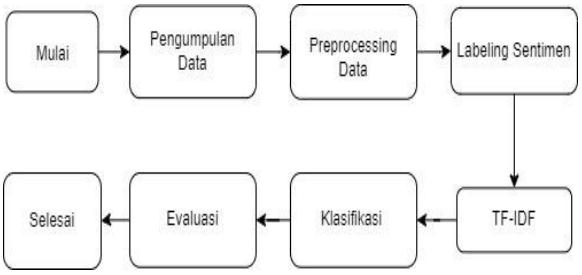
Secara umum, kasus-kasus di dunia nyata adalah kasus yang tidak linier. Untuk mengatasi sifat tidak linier dengan memakai metode kernel.Melihat bahwa dengan metode kernel suatu data x di input space dimapping ke feature space F dengan dimensi yang lebih tinggi melalui map φ sebagai berikut φ:x→φ(x). Karena itu data x di input space menjadi φ(x) di feature space. Fungsi kernel yang biasanya dipakai dalam SVM adalah Linear: xTx, Polinomial: $(x\ Txi + 1)\ p$, Radial basis function (RBF): exp $(-\ 1\ 2\sigma 2\ \|x - xi\|\ 2)$, Sigmoid: tanh $(\beta\ Txi + \beta i)$, β, $\beta i \in \Re$ [20].



Gambar 2. Kernel map mengubah persoalan yang tidak linear menjadi linier dalam space yang baru.

3. METODE PENELITIAN

Penelitian ini bertujuan untuk mengklasifikasi sentimen netizen terhadap isu pembabatan hutan adat Boven Digoel Papua dalam media sosial X melalui tagar #AllEyesOnPapua, menggunakan metode kualitatif dengan langkah-langkah yang terinci. Berikut adalah gambar alur proses penelitian:



Gambar 3. Alur Proses Penelitian

Tahap pertama penelitian adalah persiapan penelitian. Dalam tahap ini, peneliti melakukan studi pustaka untuk mendalami teori-teori yang relevan terkait budaya, nilai, dan norma yang berkembang dalam situasi sosial yang diteliti[21].

Kajian pustaka juga mencakup pemahaman mendalam tentang masalah hutan di Papua, memastikan bahwa kerangka teori yang digunakan kuat dan mendukung penelitian ini secara komprehensif. Selanjutnya, peneliti melakukan pengumpulan data training dan testing secara manual dari media sosial X. Data ini diklasifikasikan sebagai sentimen positif, negatif, atau netral, yang menjadi dasar pelatihan model[22].

Tahap kedua adalah proses penelitian yang meliputi crawling data, preprocessing data, labelling sentimen, penggunaan metode TF-IDF, pengklasifikasian menggunakan Support Vector Machine (SVM), dan evaluasi. Proses crawling data dilakukan menggunakan Google Colab untuk mengumpulkan tweet dengan tagar #AllEyesOnPapua dari media sosial X. Data kemudian melalui tahap preprocessing untuk membersihkan noise dan mempersiapkan data yang terstruktur untuk analisis lebih lanjut.

Proses evaluasi dilakukan untuk memvalidasi kinerja model SVM yang telah dilatih. Ini mencakup pengujian model dengan data testing yang tidak digunakan dalam pelatihan untuk mengukur akurasi dan kemampuan generalisasi model terhadap data baru. Evaluasi juga melibatkan penggunaan *confusion matrix* untuk mengevaluasi kinerja model klasifikasi. Matriks ini menunjukkan perbandingan antara prediksi model dan hasil sebenarnya dengan membagi data ke dalam empat kategori: *True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN)*. Matriks ini membantu dalam menghitung berbagai metrik kinerja seperti *accuracy*, *precision*, *recall*, dan *f1-score*[23].

Dengan demikian, penelitian ini tidak hanya bertujuan untuk menganalisis sentimen netizen terhadap isu pembabatan hutan adat Boven Digoel di Papua, tetapi juga untuk menunjukkan bahwa metode kualitatif, terutama dengan pendekatan berbasis teks dan analisis sentimen menggunakan SVM, dapat memberikan wawasan yang berharga dalam memahami opinis publik secara lebih mendalam melalui media sosial. Langkah-langkah yang dilakukan menjelaskan proses detail yang diperlukan untuk mencapai tujuan ini, dengan harapan hasil penelitian ini dapat memberikan kontribusi signifikan dalam konteks perlindungan lingkungan dan hak-hak masyarakat adat di Papua.

4. HASIL DAN PEMBAHASAN

Dalam penelitian ini data yang digunakan adalah tweet dengan tagar #AllEyesOnPapua dalam media sosial X sebanyak 1270 data yang diambil dari tanggal 25 Mei sampai dengan 20 Juni 2024. Setelah melewati preprocessing, tersisa 1148 data. Data-data ini kemudian di analisis untuk menemukan sentimen netizen terhadap isu pembabatan hutan di Papua. Sementara untuk data testing manual langsung diambil di media social X.

4.1. *Data Testing Manual*

Pada *data testing manual*, data diambil dari media sosial X dan dianalisis secara secara manual sebagai bahan pelatihan. Setelah melewati proses analisis dapat ditemukan tiga kategori output, yaitu sentimen positif, negatif, dan netral. Berikut adalah tabel data testing manual yang diperoleh dalam penelitian ini:

Tabel 1. *Data testing* manual

Tagar	Akun	Konten	Sentimen
#alleyesonpapua	@elsekenyap	Poster All eyes on Papua	Netral
#alleyesonpapua	@douwMoses	Poster bertulis manusia ditindas emas dikuras	Negatif
#alleyesonpapua	@langittabut08	Tweet bertuliskan “Please, please kali ini ramaikan masalah ini sampai selesai, kita bantu saudara kita di Papua, hutan hampir seluas Jakarta itu akan di babat habis. Rugi donk!	Positif

4.2. *Crawling data*

Dalam proses pengumpulan atau *crawling* data, peneliti mengumupulkan tweet dengan tagar #allayesonpapua dalam media sosial X sebanyak 1270 data yang diambil dari tanggal 25 Mei sampai dengan 20 Juni 2024.

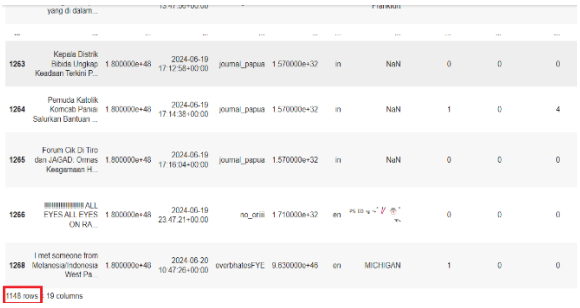
Gambar 4. Gambar hasil *crawling data*

4.3. *Preprocesing Data*

Setelah melakukan pengumpulan data, langkah selanjutnya adalah preprocessing. Hal ini dilakukan untuk menghilangkan data yang tidak relevan. Preprocessing data meliputi beberapa tahap yaitu:

a. *Cleaning* atau proses membersihkan data kesalahan dan ketidaksesuaian, seperti nilai simbol, URL, dan format yang tidak konsisten.
b. *Case Folding* atau proses mengubah semua huruf dalam teks menjadi huruf kecil untuk mengurangi variasi dalam teks dan memudahkan analisis lebih lanjut.
c. *Remove Duplicate Data* atau menghapus dataset yang sama dengan menetapkan 1 dataset utama.
d. *Tokenizing* atau proses memecah teks menjadi unit-unit kecil, yang disebut token, seperti kata, frasa, atau simbol.
e. Normalisasi atau proses menggantikan kalimat slang menjadi formal dengan menggunakan kamus colloquial-indonesian-lexicon.csv. Kamus ini berisi daftar kata-kata slang dan bentuk formalnya dalam bahasa Indonesia.
f. *Stopword Removal* atau mengurangi jumlah data yang harus diproses dan meningkatkan fokus pada kata-kata yang lebih penting atau bermakna dalam teks.
g. *Stemming* atau proses menghapus pra-pemrosesan teks untuk mengurangi kata ke bentuk dasarnya atau akar katanya.
h. *To Sentence* atau mengubah bentuk list token pada kolom stemming menjadi sentence atau kalimat.

Setelah melewati tahap-tahap di atas data yang sebelumnya 1270 tersisa 1148 data. data-data yang tidak relevan dengan penelitian ini telah dihilangkan. Data ini merupakan data yang akan digunakan untuk penelitian ini.



Gambar 5. Hasil *Preprocesing*

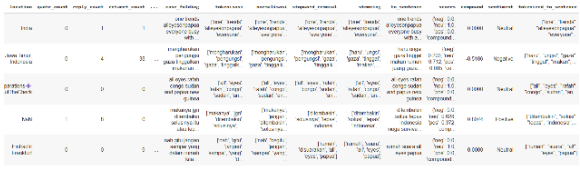
4.4. *Labeling Sentimen*

Labeling Sentimen adalah proses mengklasifikasikan teks menjadi kategori tertentu berdasarkan sentimen atau emosi yang diekspresikan dalam teks tersebut. Dalam penelitian ini, proses labeling sentiment menggunakan Library Vader Lexicon dari NLTK *(Natural Language Toolkit)*. Cara kerja library ini yakni dengan cara memberikan skor sentimen pada setiap katanya, kemudian menjumlahkannya pada setiap datanya dan menentukan hasil skor tersebut apakah bersentimen negatif, positif atau netral. Untuk catatan, library ini hanya bisa mendeteksi bahasa inggris saja, maka data sebelum diproses labelin diubah terlebih dahulu kedalam bahasa inggris menggunakan Library *googletrans*.

Gambar 6. Hasil Labeling Sentimen

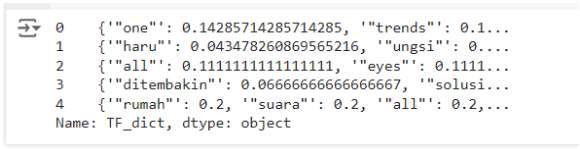
4.5. TF-IDF

TF-IDF atau *Term Frequency-Inverse Document Frequency* ini adalah tahap bertujuan untuk melihat bobot nilai dari setiap kata nya dengan menggunakan pemodelan TF-IDF. Untuk data teks yang dipakai yakni dari kolom *to_sentence* yang mana merupakan hasil akhir dari *preprocessing data*. Data teks pada kolom *to_sentence* terlebih dahulu dibuat menjadi bentuk list dengan *word_tokenized*, atau maksudnya diubah kembali dalam bentuk seperti tokenized, yang bertujuan untuk memisahkan kata per kata untuk memberikan nilai bobotnya. Untuk mencapai hasil tersebut dapat melalui beberapa langkah-langkah yang pertama yaitu melakukan proses *tokenized* kembali untuk dataset, dengan *word_tokenized* pada library NLTK. Ini merupakan proses penting ketika melakukan pembobotan pada kata di kolom *to_sentence*.



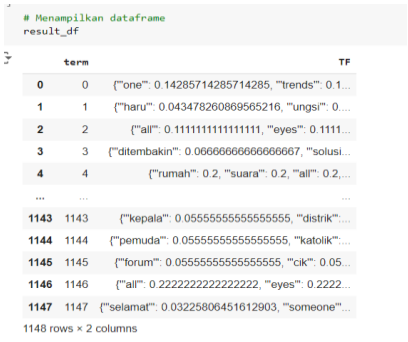
Gambar 7. Proses pembobotan kata di kolom *to_sentence.*

Langkah kedua, melakukan perhitungan bobot *(Term Frequency, TF*) dari setiap kata dalam dataset di kolom to_sentence. Hasilnya tersimpan dalam kolom 'TF_dict'.



Gambar 8. Hasil dalam kolom TF-dict

Langkah ketiga, mengambil hasil perhitungan pada kolom TF_dict atau hasil TF *(Term Frequency)* tadi dan menyajikannya kedalam bentuk dataframe/tabel agar mudah dibaca.



```
# Menampilkan dataframe
result_df
```

	term	TF
0	0	{'"one"': 0.14285714285714285, '"trends"': 0.1...
1	1	{'"haru"': 0.043478260869565216, '"ungsi"': 0....
2	2	{'"all"': 0.1111111111111111, '"eyes"': 0.1111...
3	3	{'"ditembakin"': 0.06666666666666667, '"solusi...
4	4	{'"rumah"': 0.2, '"suara"': 0.2, '"all"': 0.2,...
...	...	...
1143	1143	{'"kepala"': 0.05555555555555555, '"distrik"':...
1144	1144	{'"pemuda"': 0.05555555555555555, '"katolik"':...
1145	1145	{'"forum"': 0.05555555555555555, '"cik"': 0.05...
1146	1146	{'"all"': 0.2222222222222222, '"eyes"': 0.2222...
1147	1147	{'"selamat"': 0.03225806451612903, '"someone"'...

1148 rows × 2 columns

Gambar 9. Dataframe TF (*Term Frequency).*

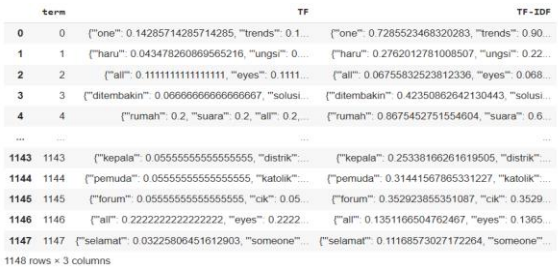
Langkah ke empat, menghitung DF (*Document Frequency*) dan menyajikannya kedalam dataframe/tabel agar mudah dibaca.

	term	DF
0	"one"	6
1	"trends"	1
2	"alleyesonpapua"	836
3	"everyone"	2
4	"busy"	1
...	...	...
3182	"user"	1
3183	"disgusted"	1
3184	"could"	1
3185	"whew"	1
3186	"wow"	1

3187 rows × 2 columns

Gambar 10. Dataframe *DF(Document Frequency).*

Langkah terakhir, menghitung bobot *Term Frequency-Inverse Document Frequency* (TF-IDF) untuk setiap kata dalam dataset dikolom Tweet.



	term	TF	TF-IDF
0	0	{'"one"': 0.14285714285714285, '"trends"': 0.1...	{'"one"': 0.728523468320283, '"trends"': 0.90...
1	1	{'"haru"': 0.043478260869565216, '"ungsi"': 0....	{'"haru"': 0.276201781008507, '"ungsi"': 0.22...
2	2	{'"all"': 0.1111111111111111, '"eyes"': 0.1111...	{'"all"': 0.06755832523812336, '"eyes"': 0.068...
3	3	{'"ditembakin"': 0.06666666666666667, '"solusi...	{'"ditembakin"': 0.4235086264213044, '"solusi...
4	4	{'"rumah"': 0.2, '"suara"': 0.2, '"all"': 0.2,...	{'"rumah"': 0.8675452751554604, '"suara"': 0.6...
...	...	...	...
1143	1143	{'"kepala"': 0.05555555555555555, '"distrik"':...	{'"kepala"': 0.25338166261619505, '"distrik"':...
1144	1144	{'"pemuda"': 0.05555555555555555, '"katolik"':...	{'"pemuda"': 0.31441567865331227, '"katolik"':...
1145	1145	{'"forum"': 0.05555555555555555, '"cik"': 0.05...	{'"forum"': 0.352923855351087, '"cik"': 0.3529...
1146	1146	{'"all"': 0.2222222222222222, '"eyes"': 0.2222...	{'"all"': 0.1351166504762467, '"eyes"': 0.1365...
1147	1147	{'"selamat"': 0.03225806451612903, '"someone"'...	{'"selamat"': 0.11168573027172264, '"someone"'...

1148 rows × 3 columns

Gambar 11. Hasil TF-IDF.

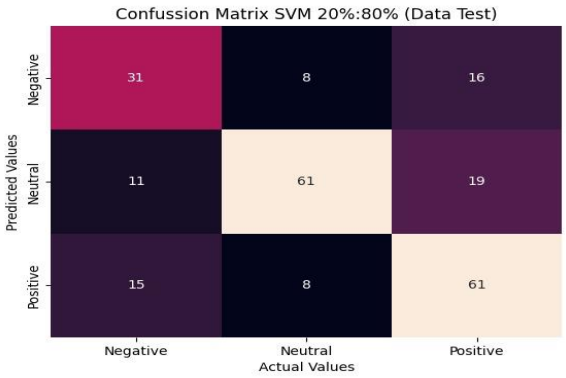
4.6. Hasil Klasifikasi dan evaluasi

Dalam penelitian ini metode *Support Vector Machine* yang dipergunakan dalam penelitian ini. Data hasil preprocessing dibagi (*spliting*) menjadi dua bagian yaitu 20% data uji dan 80% data latih sebagai data perbandingan. Sebelum mengklasifikasi data diektrasi terlebih dahulu ke nilai bobot dengan menggunakan *TfidfVectorizer* kemudian di klasifikasi menggunakan algoritma SVM dengan kernel linear dan juga memakai *class balanced* agar data yang di proses oleh sistem seimbang untuk setiap sentimennya. Kemudian untuk hasilnya berupa *classification* yang ditunjukan dalam tabel 2 :

Tabel 2. Hasil Klasifikasi

Sentiment	Precision	Recall	F1-Score
Negatif	0.56	0.54	0.55
Netral	0.61	0.79	0.73
Positif	0.73	0.64	0.68
Accuracy	0.67		

Setelah melalui proses klasifikasi untuk menilai sentimen maka diperoleh hasil 280 pendapat negatif, 394 netral, dan 474 positif. *Confusion Matrix* menunjukan performa yang didapatkan nilai *acccuracy* sebesar 67%. Pada hasil klasifikasi SVM didapat *True Negative(TN)* sebanyak 31, *True Neutral(TNR)* 61, *True Positive(TP)* 61. Pada sentimen positif didapatkan nilai *precision* sebesar 73%, *recall* 64% dan *f1-score* 68% hal ini menunjukan bahwa model cenderung cukup baik dalam mengindetifikasikan dan memprediksi sentimen positif daripada negatif dan netral.



Gambar 13 . Hasil evaluasi *Confusion matrix*

5. KESIMPULAN DAN SARAN

Dalam penelitian analisis sentimen netizen di media sosial X terhadap isu pembabatan hutan adat di Boven Digoel, Papua menggunakan metode *Support Vector Machine (SVM)*, ditemukan hasil bahwa model mencapai tingkat *accuracy* sebesar 67%, menunjukan bahwa model memprediksi dengan benar dari keseluruhan dataset. Pada sentimen Positif didapatkan *precision* sebesar 67%, *recall* 64% dan *f1-score* 68%. Hal ini menunjukan bahwa model cenderung cukup baik dalam mengindetifikasikan dan memprediksi sentimen positif daripada negatif dan netral.

DAFTAR PUSTAKA

[1] Greenpeace, *Stop Baku Tipu: Sisi Gelap Perizinan di Tanah Papua*, Amsterdam: Greenpeace International, 2021.

[2] B. Liu, "Sentiment Analysis and Opinion Mining," in *Handbook of Natural Language Processing*, 2nd ed., New York: Chapman & Hall/CRC Press, 2010, ch. Sentiment Analysis and Opinion Mining.

[3] Kanakaraj, M. & Guddeti, R.M.R. (2015). *Performance Analysis of Ensemble Methods on Twitter Sentiment Analysis Using NLP*

Techniques. Anaheim: Proc. 2015 IEEE 9th Int. Conf. Semantic Computing

[4] A. N. Ulfah and M. K. Anam, "Analisis Sentimen Hate Speech Pada Portal Berita Online Menggunakan Support Vector Machine (SVM)," *Jatisi*, vol. 7, no. 1, pp. 1-10, 2020.

[5] H. C. Husada and A. S. Paramita, "Analisis Sentimen Pada Maskapai Penerbangan di Platform Twitter Menggunakan Algoritma Support Vector Machine (SVM)," *TEKNIKA*, vol. 10, no. 1, pp. 18-26, 2021.

[6] B. Pamungkas, M. Eka Purbaya, and D. A. K. Januarita, "Analisis Sentimen Twitter Menggunakan Metode Support Vector Machine (SVM) pada Kasus Benih Lobster 2020," *Journal of INISTA*, vol. 3, no. 2, pp. 10-20, 2021.

[7] Que, V. K. S., Iriani, A., & Purnomo, H. D. (2020). Analisis Sentimen Transportasi Online Menggunakan Support Vector Machine Berbasis Particle Swarm Optimization. *Jurnal Nasional Teknik Elektro dan Teknologi Informasi*, 9 (2), 162- 170

[8] B. Liu, "Sentiment Analysis and Opinion Mining," in *Handbook of Natural Language Processing*, 2nd ed., New York: Chapman & Hall/CRC Press, 2010, ch. Sentiment Analysis and Opinion Mining.

[9] P. D. Turney, "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews," in *Association for Computational Linguistics 40th Anniversary Meeting*, New Brunswick, N.J., 2002.

[10] B. Pang, L. Lee, and S. Vaithyanathan, "Sentiment Classification Using Machine Learning Techniques," in *Proc. 7th Conf. Empirical Methods in Natural Language Processing (EMNLP-02)*, 2002, pp. 79-86.

[11] A. Nurzahputra and M. A. Muslim, "Analisis Sentimen pada Opini Mahasiswa Menggunakan Natural Language Processing," in *Seminar Nasional Ilmu Komputer (SNIK 2016)*, Semarang, 2016.

[12] D. Jurafsky and H. J. Martin, *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, New Jersey: Prentice-Hall, 2007.

[13] S. Nurwanda, N. Suarna, and W. Prihartono, "Penerapan NLP (Natural Language Processing) Dalam Analisis Sentimen Pengguna Telegram Di Playstore," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 2, pp. 1841-1846, 2024.

[14] W. I. F. Ali Firdaus, "Text Mining Dan Pola Algoritma Dalam Penyelesaian Masalah Informasi: (Sebuah Ulasan)," *Jurnal Politeknik Negri Sriwijaya*, vol. XIII, p. 66, 2021.

[15] F.-C. Lee, "Comparison of the Primitive Classifiers without Features Selection in Credit Scoring," in *Management and Service Science*, 2009.

[16] O. Maimon and L. Rokach, *Data Mining and Knowledge Discovery Handbook*, New York: Springer, 2005.

[17] P. P. Widodo, R. T. Handayanto, and Herlawati, *Penerapan Data Mining Dengan Matlab*, Bandung: Rekayasa Sains, 2013.

[18] B. Santosa, *Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis*, Yogyakarta: Graha Ilmu, 2007.

[19] A. S. Ritonga and E. S. Purwaningsih, "Penerapan Metode Support Vector Machine (SVM) Dalam Klasifikasi Kualitas Pengelasan SMAW (Shield Metal Arc Welding)," *Jurnal Ilmiah Edutic*, vol. 5, no. 1, pp. 17-25, 2018.

[20] V. N. Vapnik, *The Nature of Statistical Learning Theory*, 2nd ed., New York: Springer-Verlag, 1999.

[21] Sugiyono, *Metode Penelitian Kuantitatif Kualitatif dan R&D*, Bandung: Alfabeta, 2012.

[22] C. F. Hasri and D. Alita, "Penerapan Metode Naïve Bayes Classifier Dan Support Vector Machine Pada Analisis Sentimen Terhadap Dampak Virus Corona Di Twitter," *Jurnal Informatika dan Rekayasa Perangkat Lunak (JATIKA)*, vol. 3, no. 2, pp. 145-160, 2022.

[23] J. Davis and M. Goadrich, "The relationship between Precision-Recall and ROC curves," in *Proceedings of the 23rd International Conference on Machine Learning (ICML)*, 2006, pp. 233-240.