Jurnal Ilmu Komputer dan Informatika (JIKI)
Vol. 4, No. 1, Juni 2024, Hal. 43-50
https://jiki.jurnal-id.com
DOI: https://doi.org/10.54082/jiki.141
P-ISSN: 2807-6664
E-ISSN: 2807-6591

Analisis Sentimen Masyarakat terhadap Kasus Korupsi PT. Timah Menggunakan Metode Support Vector Machine

Fionna Caroline[*1], Raden George Samuel Budi[2], Muhammad Ezar Al Rivan[3]

[1,2,3]Informatika, Fakultas Ilmu Komputer dan Rekayasa, Universitas Multi Data Palembang, Indonesia
Email: [1]fionnacaroline@mhs.mdp.ac.id, [2]radengeorge@mhs.mdp.ac.id,
[3]meedzhar@mdp.ac.id

Abstrak

Korupsi adalah penyalahgunaan jabatan publik untuk keuntungan pribadi yang dimana korupsi ini dapat memberikan kerugian besar bagi negara maupun masyarakat. Topik yang dipilih untuk penelitian ini adalah kasus korupsi PT. Timah yang sedang hangat dibicarakan dikarenakan kerugian negara yang mencapai 271 T. Untuk membantu analisis dalam penelitian ini, dibangunlah sebuah sistem yang dapat mendeteksi sentimen publik yang sudah dikumpulkan dari platform Youtube dengan *metode Support Vector Machine*. Model yang sudah dilatih dengan dataset akan diseimbangkan dengan SMOTE karena tidak meratanya kelas klasifikasi. Model klasifikasi yang telah dibangun dengan support vektor machine mendapatkan hasil presisi pada sentimen negatif 91% dan sentimen positif 44%, *recall* pada sentimen negatif 96% dan sentimen positif 22%, F1-Score pada sentimen negatif 93% dan sentimen positif 30%, serta jumlah *sample* pada kelas sentimen negatif 140 dan kelas sentimen positif 18.

Kata kunci: *Klasifikasi, Korupsi, Sentimen, SMOTE, Support Vector Machine (SVM)*

Abstract

Corruption is the abuse of public office for personal gain, where this corruption can cause major losses to the state and society. The topic chosen for this research is the PT.Timah corruption case where currently hotly discussed due to state losses reaching 271 T. To assist the analysis in this research, a system was built that can detect public sentiment which has been collected from the YouTube platform using the Support Vector Machine method. Models that have been trained with the dataset will be balanced with SMOTE because of the uneven classification classes. The classification model that has been built with a support vector machine gets precision results on negative sentiment of 91% and positive sentiment of 44%, recall on negative sentiment of 96% and positive sentiment of 22%, F1-Score on negative sentiment of 93% and positive sentiment of 30%, and the number of samples in the negative sentiment class is 140 and the positive sentiment class is 18.

Keywords*: classification, corruption, sentiment, SMOTE, support vector machine*

1. PENDAHULUAN

Korupsi adalah penyalahgunaan jabatan publik untuk keuntungan pribadi. Segala bentuk pemerintahan/badan pemerintahan sebenarnya rentan terhadap korupsi. Perbuatan korupsi tersebut dilakukan dengan tujuan untuk memberikan keuntungan yang tidak sesuai dengan tugas atau hak masyarakat, dapat berupa dana atau barang negara, untuk memperkaya diri sendiri [1]. Kasus korupsi terbaru yang banyak menggiring berbagai opini dari masyarakat adalah kasus korupsi penambangan timah ilegal PT. Timah (Tbk) dari tahun 2015 hingga 2022, negara mengalami kerugian sebesar 271 069.688.018.700. Dari kerugian tersebut, kerugian lingkungan (ekologis) sebesar 157.832.395.501.025, kerugian ekonomi lingkungan sebesar 60.276.600.800.000, dan biaya pemulihan lingkungan sebesar Rp. 6.257.249.726.025. Selain itu terdapat juga kerugian di luar kawasan hutan sekitar Rp. 47.703.441.991.650 [2].

Analisis sentimen adalah teknik yang mengekstrak data opini dan secara otomatis memahami serta memproses data teks untuk mengidentifikasi sentimen dalam opini [3]. Analisis sentimen berfokus pada

DOI: https://doi.org/10.54082/jiki.141
P-ISSN: 2807-6664
E-ISSN: 2807-6591

pengolahan opini yang menunjukkan polaritas, yaitu nilai sentimen positif atau negatif [4]. Analisis sentimen saat ini dibagi menjadi penggunaan metode klasifikasi dan metode klasifikasi pembelajaran mesin berbasis aturan. Teknik pembelajaran mesin menggunakan kata-kata emosi sebagai fitur klasifikasi, dan pemilihan emosi dapat dicapai dengan cepat dan efisien menggunakan kamus. Biasanya, tugas yang melibatkan klasifikasi menggabungkan penggunaan teknik *Naive Bayes, Support Vector Machine* (SVM), dan *Maximum Entropy* [5]

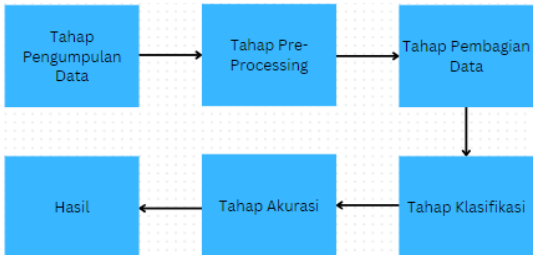
Penelitian mengenai analisis sentimen sebelumnya sudah banyak dilakukan, seperti Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode *Support Vector Machine* [6], penelitian ini dilakukan perbandingan dua kernel SVM, yaitu *polynomial* dan *RBF*. Hasil yang diperoleh dari pengujian yang dilakukan kernel *polynomial* menghasilkan akurasi 98.67%, sedangkan kernel *RBF* menghasilkan akurasi 98.34%. Selain itu, penelitian yang berjudul Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes [7], mendapatkan hasil akurasi 77.78%. Penelitian berjudul *Sentiment Analysis of YouTube Movie Trailer Comments Using Naïve Bayes* [8] dan mendapatkan hasil akurasi sebesar 81%.

Beberapa penelitian menggunakan dataset yang tidak seimbang, maka dataset tersebut diseimbangkan terlebih dahulu, seperti pada penelitian yang menganalisis sentimen terhadap aplikasi Ruangguru dengan menggunakan beberapa model algoritma, yaitu *Naïve Bayes, Random Forest dan Support Vector Machine* [9]. Penelitian ini menggunakan metode SMOTE untuk menyeimbangkan data. Hasil yang didapat dalam pengujian yang dilakukan pada penellitian ini yaitu, model algoritma *Random Forest* memiliki nilai akurasi tertinggi dengan 97,16% dan nilai AUC 0,996. Algoritma *Support Vector Machine* memiliki akurasi sebesar 96,01% dengan nilai AUC 0,543, dan algoritma *Naive Bayes* memiliki akurasi terendah dengan 94,16% dan nilai AUC 0,999.

Berdasarkan permasalahan diatas, penelitian ini dilakukan menggunakan metode *SVM* sebagai model klasifikasi. Namun dataset yang digunakan pada penelitian ini tidak seimbang. Mengingat nilai presisi tinggi tidak hanya didasarkan pada algoritma, tetapi juga pada karakteristik dataset, maka ketidakseimbangan dalam dataset perlu diatasi. Ketidakseimbangan data menyebabkan performa klasifikasi tidak optimal, karena model belajar dari lebih banyak data masukan daripada kelas mayoritas [10]. Oleh karena itu, penelitian ini mengunakan metode SMOTE untuk mengatasi ketidakseimbangan dataset. Synthetic Minority Oversampling Technique (SMOTE) merupakan teknik resampling yang bertujuan untuk menyeimbangkan distribusi kelas. Untuk menambah jumlah data pada kelas minoritas, seseorang harus mengambil sampel dari kelompoknya dan memasukkan sampel sintetik [10]

2. METODE PENELITIAN

Penelitian ini dimulai dengan pengumpulan data dari komentar pada video YouTube terkait kasus korupsi PT. Timah, menghasilkan total 1186 komentar yang kemudian disimpan untuk analisis lebih lanjut. Tahap selanjutnya adalah *preprocessing* data, yang mencakup pembersihan data dengan menghilangkan karakter khusus, URL, dan elemen tidak relevan, serta tokenisasi dan stemming untuk mempersiapkan data. Setelah itu, metode SMOTE diterapkan untuk mengatasi ketidakseimbangan data dengan menambah jumlah data positif sehingga distribusi menjadi lebih seimbang.



Gambar 1. Tahapan Penelitian

Data yang telah dipreproses dan diimbangi kemudian dibagi menjadi data latih dan data uji. Model SVM dilatih menggunakan data latih untuk mengklasifikasikan sentimen menjadi positif dan negatif, dan kemudian dievaluasi menggunakan data uji dengan metrik precision, recall, F1-Score, dan akurasi.

DOI: https://doi.org/10.54082/jiki.141
P-ISSN: 2807-6664
E-ISSN: 2807-6591

Hasil evaluasi menunjukkan bahwa model SVM memiliki performa yang baik dalam mengklasifikasikan sentimen negatif tetapi kurang efektif dalam mendeteksi sentimen positif, menunjukkan perlunya peningkatan teknik pengimbangan data atau eksplorasi metode klasifikasi lain untuk hasil yang lebih baik

2.1. Metode Support Vector Machine

Support Vector Machine (SVM) adalah suatu teknik untuk melakukan prediksi, baik dalam kasus klasifikasi maupun regresi. *SVM* memiliki prinsip dasar *linear classifier* yaitu kasus klasifikasi yang secara *linier* dapat dipisahkan, namun SVM telah dikembangkan agar dapat bekerja pada problem *non-linier* dengan memasukkan konsep kernel pada ruang kerja berdimensi tinggi. Pada dasarnya, metode ini bekerja dengan cara mendefinisikan batas antara dua kelas dengan jarak maksimal dari data yang terdekat. Untuk mendapatkan batas maksimal antar kelas maka harus dibentuk sebuah hyperplane (garis pemisah) terbaik pada input space yang diperoleh dengan mengukur margin hyperplane dan mencari titik maksimalnya [11]. Klasifikasi dilakukan dengan mencari *hyperplane* atau garis pembatas (*decision boundary*) yang memisahkan antara suatu kelas dengan kelas lain, *Support Vector Machin*e melakukan pencarian nilai *hyperplane* dengan menggunakan *support vector* dan nilai *margin*[9].

2.2. Pengumpulan Data

Dataset yang digunakan dalam penelitian ini adalah data primer. Dataset berisi teks berbahasa Indonesia yang diperoleh dari komentar pada aplikasi youtube dengan tema korupsi PT. Timah (link : https://www.youtube.com/watch?v=Jlk1ljeGdJE&t=25s). Data yang diambil sebanyak 1186 komentar. Pengumpulan data ini menggunakan sebuah web yang bernama "*Netlytic*" untuk mengambil sentimen-sentimen yang terdapat pada video youtube tersebut. Hasil pengumpulan data dapat dilihat pada table 1.

Tabel 1. Contoh Sampel Data

Id	Author	Pubdate	Title
1	@nagabettafarm4729	2024-05-08 0:19:41	Jenius loh sih HM ini, saking jenius nya jd salah jalan
2	@purnomosh4858	2024-05-07 20:10:12	Korup😊r harus dimiskinkan termasuk perampasan harta milik tapi apa UU nya sdh disetuji DPR ?
3	@edoedward	2024-05-06 20:07:05	Really appreciate how you break things down to educate public. Subscribing!
4	@wicengvlog8489	2024-05-06 19:14:48	Enk ya krja nya korupsi.. dpat 50T hukum paling 5thun. itung itung d hukum 5tahun itu kita krja dpat 50T kotor lah.
5	@wongdeso9439	2024-05-06 6:58:44	Saya subcribe bang detil banget penjelasan nya

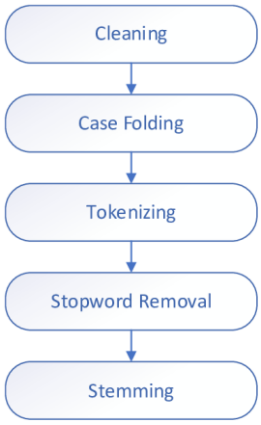
2.3. Pra-proses

Pada tahap pra-proses, data yang digunakan peneliti adalah data yang hanya diperlukan saja. Yang pertama akan dilakukan *read data .csv*, data ini didapatkan sesudah melewati proses dan data tersebut akan memiliki format *.csv*, data ini akan diproses menggunakan pemrograman bahasa Python dengan menggunakan *library Pandas* agar data .csv tersebut dapat di baca. Lalu akan dilakukan Case Folding, yaitu data tersebut akan diubah semuanya menjadi "*lowercase*".

Data Cleaning, dengan tahapan pertama adalah *Tokenizing*, yaitu teks yang diperoleh dari data akan dipisah untuk tiap katanya dan mungkin pada waktu bersamaan akan dihapuskan karakter tertentu, lalu *Filtering*, mengeliminasi kata yang tidak memiliki pengaruh atau tidak memiliki informatif [12], *Spelling* memperbaiki kata-kata yang memiliki kesalahan dalam penulisan dan kata-kata yang disingkat, *N-gram* menghubungkan kata-kata yang memiliki keterkaitan. Hasil dari proses *pre-processing* tersebut

DOI: https://doi.org/10.54082/jiki.141
P-ISSN: 2807-6664
E-ISSN: 2807-6591

berupa data yang berkualitas yang dapat mempermudah dalam proses klasifikasi [13]. Alur pra-proses yang dilakukan dapat dilihat pada Gambar 2.



Gambar 2. Alur Pra-proses

2.3.1. Tahap Pembagian Data

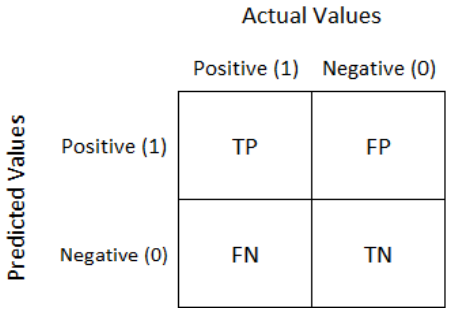
Kumpulan kata yang di dapat di data memiliki ekspresi sentimen yang berupa positif, negatif dan netral. *Leksikon* sentimen dibentuk dengan cara manual dengan cara mengelompokkan atau labelling kata-kata sentimen menjadi sentimen positif, negatif dan netral secara manual dilakukan oleh peneliti [14]. Data yang digunakan akan menjadi data training dan data testing.

2.3.2. Tahap Ekstraksi Fitur

Dalam menganalisa sentimen, perlu dilakukan ekstraksi fitur fitur yang relevan untuk melihat *sentiment analysis* yang dipergunakan untuk melihat melihat pendapat atau kesamaan terhadap suatu persoalan atau objek oleh seorang menuju opini yang positif atau negatif. Dalam metode *TF-IDF*, Frekuensi Istilah lebih menitikberatkan pada istilah yang sering muncul dalam satu dokumen, sedangkan Frekuensi Dokumen Terbalik lebih mengutamakan pemberian bobot rendah untuk istilah yang sering muncul di banyak dokumen [15]. Maka dari itu diterapkannya TF-IDF untuk memperhitungkan frekuensi setiap sentimen untuk mendapatkan bobot yang diperlukan untuk tahap klasifikasi.

2.3.3. Tahap Klasifikasi

Data yang telah diproses dan ekstraksi fiturnya akan diklasifikasikan menggunakan metode *SVM*. Klasifikasi ini bertujuan untuk mencari keputusan yang terbaik ke dalam sentimen positif dan negatif, untuk selanjutnya data akan di *training*. Banyak penelitian telah menunjukkan bahwa *SVM* adalah metode yang paling akurat untuk klasifikasi teks. Performa model penelitian akan diukur dalam c*onfusion matrix* seperti pada Gambar 4. *Confusion Matrix* adalah pengukuran performa untuk masalah klasifikasi *machine learning* dimana keluaran dapat berupa dua kelas atau lebih. *Confusion Matrix* adalah tabel dengan 4 kombinasi berbeda dari nilai prediksi dan nilai aktual [10].



Gambar 4. Confusion Matrix

DOI: https://doi.org/10.54082/jiki.141
P-ISSN: 2807-6664
E-ISSN: 2807-6591

Untuk menghitung hasil kinerja model berdasarkan informasi dalam *confusion matrix*, dapat menggunakan rumus sebagai berikut :

$$precision = \frac{TP}{TP+FP} \quad (1)$$

$$recall = \frac{TP}{TP+FN} \quad (2)$$

$$F1 = \frac{2 \times precision \times recall}{precision+recall} \quad (3)$$

$$accuracy = \frac{TP+TN}{TP+FN+TN+FP} \quad (4)$$

$$specicity = \frac{TN}{TN+FP} \quad (5)$$

3. HASIL DAN PEMBAHASAN

3.1. *Preprocessing* Dataset

Diperoleh dataset melalui dari sentimen-sentimen yang terdapat pada komentar youtube mengenai kasus korupsi 271 Trilliun (https://www.youtube.com/watch?v=Jlk1ljeGdJE&t=25s), pada tahap *scrapping* di dapatkan 1186 komentar data dengan rentang waktu awal mulai video di upload sampai pada 15 Maret 2024. Dan dilakukan text *preprocessing* pada dataset tersebut. Seperti pengecilan kapital huruf, penghapusan *stopword*, *normalisasi*, *tokenisasi*, dan dilakukan *stemming*. Data kemudian disimpan di file terpisah dan hasil *preprocessing* dapat dilihat pada gambar 5.

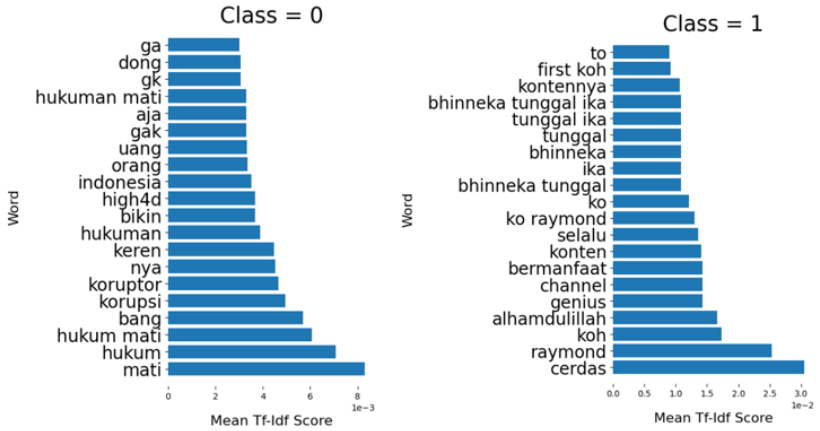
Gambar 5. Visualisasi hasil data *preprocessing*

Seperti yang sudah dilihat pada Gambar 5, dataset yang sudah di-*praprocessing* akan ditampilkan dalam bentuk *word cloud* yang ditampilkan dalam bentuk token di setiap pengelompokannya yang dimana kata "koruptor" muncul di kedua kelompok data kata yang merupakan kata itu sendiri adalah kata pokok atau utama dalam penelitian ini.

3.2. Penerapan Ekstraksi Fitur

Setelah dilakukannya *preprocessing data* yang dimana kata, dilakukannya ekstraksi fitur menggunakan metode TF-IDF . Dapat dilihat pada gambar ?, yang merupakan hasil dari ekstraksi fitur menggunakan TF-IDF yang akan digunakan untuk klasifikasi nanti.

DOI: https://doi.org/10.54082/jiki.141
P-ISSN: 2807-6664
E-ISSN: 2807-6591

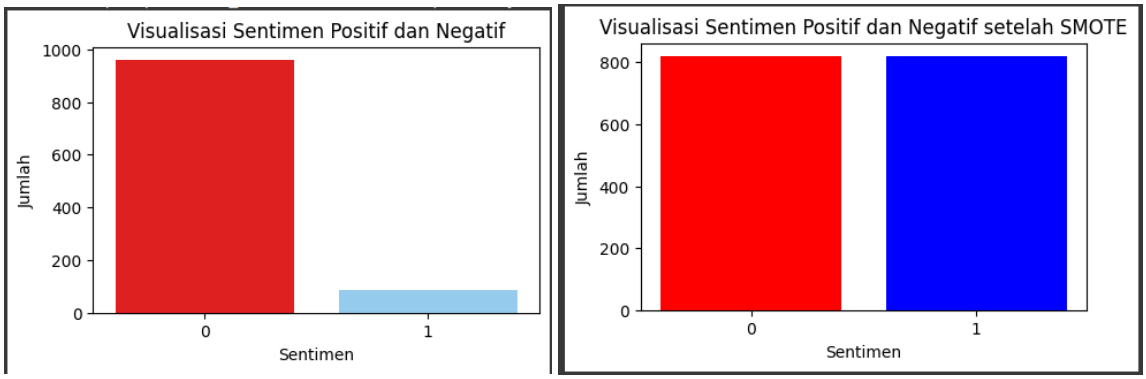


Gambar 6. Hasil Ektraksi Fitur

Dapat dilihat pada Gambar 6, Yang dimana setiap kata sudah memiliki bobotnya sendiri. Dan juga Pembobotan TF-IDF ini dilakukan agar pada tahap menentukan klasifikasi sentimen terhadap topik korupsi di PT. Timah dapat dilakukan.

3.3. Penerapan SMOTE

Setelah dilakukannya ekstraksi fitur, dikarenakan terdapatnya ketidakseimbangan dataset yang digunakan, oleh karena itu untuk menyeimbangkan dataset, peneliti menggunakan metode *SMOTE* agar pelabelan antara sentimen positif dan sentimen negatif bisa diseimbangkan.



Gambar 7. Menyeimbangkan Dataset

Pembobotan SMOTE ini dilakukan agar antara sentimen positif dan sentimen negatif bisa disetarakan. Seperti pada Gambar 7 yang dimana sebelum menggunakan SMOTE, perbandingan data antara sentimen positif dan negatif mencapai 3:1, setelah dilakukannya SMOTE, rasio data menapai 1:1 antara sentimen negatif dan sentimen positif.

3.4. Pengujian Klasifikasi

Pengujian ini menggunakan metode klasifikasi SVM. Dalam penelitian ini seluruh tahapan uji coba menggunakan klasifikasi pada dataset yang telah diolah menggunakan *Python* dengan l*ibrary Pandas*.

Tabel 2. Laporan Klasifikasi Sebelum Smote

	Precision	**Recall**	**F1-Score**	**Support**
Negatif	0.89	1.00	0.94	140
Positif	0.00	0.00	0.00	18
Accuracy	-	-	0.89	158

DOI: https://doi.org/10.54082/jiki.141
P-ISSN: 2807-6664
E-ISSN: 2807-6591

Pada tabel 2 diketahui bahwa score yang dimiliki oleh Negatif lebih besar dibandingkan dengan positif ini menunjukkan bahwa dari hasil analisis dan pengujian menggunakan dataset tersebut didapatkan banyak sentimen-sentimen yang negatif dibandingkan dengan sentimen yang positif. Dapat dilihat pada table 2 yang menunjukkan nilai Precision pada negatif 89% dan positif 0% , *Recall* pada negatif 100% dan positif 0%, *F1-Score* pada nilai negatif 94% dan positif 0%, dan *Support sample* pada nilai negatif 140 dan positif 18.

Tabel 3. Laporan Klasifikasi dengan Smote

	Precision	**Recall**	**F1-Score**	**Support**
Negatif	0.91	0.96	0.93	140
Positif	0.44	0.22	0.30	18
Accuracy	-	-	0.88	158

Pada tabel 3 dapat diketahui bahwa *score* yang dimiliki oleh negatif lebih besar dibandingkan dengan positif, hasil ini menunjukkan bahwa dari hasil analisis dan pengujian menggunakan dataset tersebut didapatkan banyak sentimen-sentimen yang negatif dibandingkan dengan sentimen yang positif. Dapat dilihat pada gambar 8 yang menunjukkan nilai *Precision*, *Recall*, *F1-Score*, dan *Support sample*. Selain itu, penelitian yang dilakukan dengan menggunakan metode *SVM* mendapat akurasi cukup tinggi yaitu 88%. Bisa kita lihat dari hasil klaisifikasi dari tabel 3, dengan akurasi 88% yang dimana metode *support vector machine* dengan penyeimbang dataset menggunakan SMOTE dapat dikatakan sudah baik.

4. KESIMPULAN

Berdasarkan hasil penelitian yang ditampilkan pada tabel 3, maka disimpulkan bahwa analisis sentimen masyarakat terhadap kasus korupsi PT. Timah menggunakan metode *SVM* didapatkan akurasi sebesar 88%. Penelitian ini mengambil hasil menggunakan metode *SMOTE*, karena dengan *SMOTE* didapatkan sebuah hasil yang seimbang dibandingan dengan hasil tanpa *SMOTE*. Hasil yang akurasi yang didapatkan sebelum *SMOTE* adalah presisi pada negatif 89% dan positif 0%, *Recall* pada negatif 100% dan positif 0%, *F1-Score* pada nilai negatif 94% dan positif 0%, dan *Support* pada nilai negatif 140 dan positif 18.

Sementara hasil menggunakan SMOTE adalah Precision pada negatif 91% dan positif 44% , Recall pada negatif 96% dan positif 22%, F1-Score pada nilai negatif 93% dan positif 30%, dan Support pada nilai negatif 140 dan positif 18. Penelitian ini menggunakan data komentar masyarakat pada salah satu video youtube dengan total 1186 komentar. Data yang didapat sebagian besar merupakan komentar negatif, maka dapat disimpulkan bahwa sentimen masyarakat terhadap kasus korupsi PT. Timah cenderung negatif yang dapat dilihat pada nilai pada klasifikasi lebih condong ke negatif.

DAFTAR PUSTAKA

[1] T. Bukhary, J. Pendidikan, A. dan Sains, and D. Putri, “Korupsi dan Perilaku Koruptif,” 2021.

[2] Z. Hanyfah, A. Oktapia, and M. Tirta, “Analisis Perhitungan Kerugian Negara dari Hasil Dugaan Tindak Pidana Korupsi yang Dilakukan oleh PT. Timah (Tbk),” *Journal of Law and Nation (JOLN)*, vol. 3, no. Mei, pp. 351–358, 2024.

[3] F. V. Sari and A. Wibowo, “Analisis Sentimen Pelanggan Toko Online JD. ID Menggunakan Metode Naive Bayes Classifier Berbasis Konversi Ikon Emosi,” *Jurnal SIMETRIS*, vol. 10, no. 2, 2019.

[4] A. Novantirani, M. S. Kania Sabariah, and V. Effendy, “Analisis Sentimen pada Twitter untuk Mengenai Penggunaan Transportasi Umum Darat Dalam Kota dengan Metode Support Vector Machine,” 2015.

[5] A. Rahman Isnain, A. Indra Sakti, D. Alita, and N. Satya Marga, “Sentimen Analisis Publik Terhadap Kebijakan Lockdown Pemerintah Jakarta Menggunakan Algoritma SVM,” *JDMSI*, vol. 2, no. 1, pp. 31–37, 2021, [Online]. Available: https://t.co/NfhnfMjtXw

DOI: https://doi.org/10.54082/jiki.141
P-ISSN: 2807-6664
E-ISSN: 2807-6591

[6] N. Hendrastuty, A. Rahman Isnain, and A. Yanti Rahmadhani, "Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode Support Vector Machine," vol. 6, no. 3, 2021, [Online]. Available: http://situs.com

[7] B. Gunawan, H. Sasty, P. #2, E. Esyudha, and P. #3, "JEPIN (Jurnal Edukasi dan Penelitian Informatika) Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes," vol. 4, no. 2, pp. 17–29, 2018, [Online]. Available: www.femaledaily.com

[8] R. Novendri, A. S. Callista, D. N. Pratama, and C. E. Puspita, "Sentiment Analysis of YouTube Movie Trailer Comments Using Naïve Bayes," *Bulletin of Computer Science and Electrical Engineering*, vol. 1, no. 1, pp. 26–32, Jun. 2020, doi: 10.25008/bcsee.v1i1.5.

[9] E. Fitri, Y. Yuliani, S. Rosyida, and W. Gata, "Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma Naive Bayes, Random Forest Dan Support Vector Machine," *TRANSFORMTIKA*, vol. 18, no. 1, pp. 71–80, 2020, [Online]. Available: www.nusamandiri.ac.id,

[10] K. Pramayasa, I. Md, D. Maysanjaya, G. Ayu, and A. Diatri Indradewi, "Analisis Sentimen Program Mbkm Pada Media Sosial Twitter Menggunakan KNN Dan SMOTE", [Online]. Available: https://doi.org/10.31598

[11] R. Achmad Rizal, I. Sanjaya Girsang, and S. Apriyadi Prasetiyo, "Klasifikasi Wajah Menggunakan Support Vector Machine (SVM)," *Riset dan E-Jurnal Manajemen Informatika Komputer*, vol. 3, no. 2, 2019.

[12] F. Amin, "Sistem Temu Kembali Informasi dengan Metode Vector Space Model," 2012.

[13] J. Han, M. Kamber, and J. Pei, "Data Mining. Concepts and Techniques, 3rd Edition (The Morgan Kaufmann Series in Data Management Systems)," 2011.

[14] A. I. Tanggraeni and M. N. N. Sitokdana, "Analisis Sentimen Aplikasi E-Government Pada Google Play Menggunakan Algoritma Naïve Bayes," vol. 9, no. 2, pp. 785–795, 2022.

[15] D. Saputra and M. R. Pribadi, "Analisis Sentimen Masyarakat Terhadap Layanan Provider Internet di Indonesia Menggunakan SVM," *2 ND MDP STUDENT CONFERENCE (MSC) 2023*, 2023.