

Opini Publik Terhadap Liga Korupsi Indonesia Pada Platform Youtube Menggunakan Naïve Bayes dan SMOTE

Aldi Ardiansyah¹, Mecha Bella Permata Sihombing² and Nur Rachmat^{3,*}

¹ Program Studi Informatika Fakultas Ilmu Komputer dan Rekayasa Universitas Multi Data Palembang
aldiardiasnyah_2226250059@mhs.mdp.ac.id; mechabellapermatasihombing_2226250095@mhs.mdp.ac.id;
rachmat.nur91@mdp.ac.id

* Korespondensi : aldiardiasnyah_2226250059@mhs.mdp.ac.id

Info Artikel:

Dikirim: 09 Juni 2025

Direvisi: 14 Juli 2025

Diterima: 13 November 2025

Abstract: This study analyzes public opinion regarding the Corruption League in Indonesia by utilizing the Naïve Bayes method combined with the Synthetic Minority Oversampling Technique (SMOTE). The Corruption League is a compilation of corruption cases involving public officials, politicians, and other parties in Indonesia. In this research, Naïve Bayes is employed for sentiment classification, while SMOTE is used to address class imbalance within the dataset, which was collected from YouTube comments. The methodology consists of several stages, including data collection, labeling, preprocessing, classification, and model evaluation. The results reveal that Naïve Bayes without SMOTE achieves high performance in identifying the negative class but struggles significantly in recognizing the positive class, leading to an imbalanced classification outcome. Conversely, when Naïve Bayes is combined with SMOTE, the model's performance becomes more balanced, showing a notable improvement in detecting the positive class. Additionally, accuracy increases from 79.7% to 84.3%. This study provides valuable insights into public perceptions and demonstrates the effectiveness of classification methods in the context of corruption issues in Indonesia.

Keywords: Corruption League; NB; Public Opinion; Synthetic Minority Oversampling Techniqu

Intisari: Penelitian ini menganalisis opini publik terhadap liga korupsi di Indonesia menggunakan metode Naïve Bayes dan Synthetic Minority Oversampling Technique (SMOTE). Liga korupsi adalah Liga korupsi indonesia adalah daftar kasus korupsi yang melibatkan pejabat publik, politisi, atau pihak-pihak lain di indonesia. Dalam penelitian ini Naïve Bayes diterapkan untuk klasifikasi sentimen, sedangkan SMOTE diterapkan untuk mengatasi ketidakseimbangan kelas dalam dataset yang diambil melalui komentar Youtube. Metodologi mencakup tahapan pengumpulan data, pelabelan, preprocessing, klasifikasi, dan evaluasi model. Hasil menunjukkan bahwa Naïve Bayes tanpa SMOTE menghasilkan performa yang tinggi pada kelas negatif namun sangat lemah dalam mengenali kelas positif sehingga model tidak seimbang dalam melakukan klasifikasi. Sebaliknya penggunaan Naïve Bayes dengan menggunakan SMOTE, performa model menjadi lebih seimbang dengan peningkatan yang signifikan pada kelas positif. Nilai akurasi juga mengalami kenaikan dari 79,7% menjadi 84,3%. Dari Penelitian ini memberikan wawasan penting tentang persepsi publik dan efektivitas metode klasifikasi dalam konteks isu korupsi di Indonesia.

Kata Kunci: Liga Korupsi; NB; Opini Publik; Synthetic Minority Oversampling Technique

1. Pendahuluan

Korupsi dari bahas latin corruptus dan corruption, yang memiliki arti rusak, bejat, menyimpang dari nilai-nilai kesucian, serta tindakan yang menghina dan memfitnah. dalam modul Komisi Pemberantasan Korupsi, korupsi diartikan sebagai tindakan yang dilakukan dengan kewajiban resmi dan prinsip-prinsip kebenaran. Tindakan ini termasuk penyalahgunaan jabatan yang melanggar hukum demi keuntungan diri sendiri yang pada dasarnya bertolak belakang dengan tanggung jawab serta nilai kejujuran[1]. Tindakan Korupsi sangat merugikan negara karena dapat menghambat pertumbuhan ekonomi negara, menurunnya investasi, meningkatnya kemiskinan serta menurunnya pendapatan pada suatu negara. Tindakan ini biasanya bertujuan untuk memperoleh keuntungan yang tidak sah, baik berupa dana maupun aset negara pada akhirnya memperkaya pelaku yang melakukan korupsi secara pribadi.

Korupsi telah menjadi salah satu masalah kronis yang mengakar dalam sistem pemerintahan Indonesia. Fenomena "Liga Korupsi Indonesia" yang menjadi viral di media sosial menggambarkan keprihatinan masyarakat terhadap maraknya kasus korupsi yang melibatkan berbagai pejabat publik. Istilah ini mencerminkan persepsi publik bahwa korupsi di Indonesia telah mencapai level yang sangat mengkhawatirkan, seolah-olah para koruptor membentuk suatu "liga" atau organisasi terstruktur.

Salah satu cara untuk menggali opini publik terhadap isu ini adalah melalui analisis sentimen yang merupakan proses mengekstraksi, memahami, dan mengklasifikasikan opini atau perasaan dari suatu teks. Media sosial, seperti YouTube, sering dijadikan tempat masyarakat dalam menyuarakan opini mereka dengan berbagai permasalahan, termasuk Liga Korupsi Indonesia. Liga korupsi indonesia adalah daftar kasus korupsi yang melibatkan pejabat publik, politisi, atau pihak-pihak lain di indonesia, dikutip dari kompas.com kerugian mencapai hampir 1 kuadriliun[2].

Metode yang efektif dalam menganalisis opini publik adalah algoritma Naïve Bayes dan SMOTE. Algoritma Naïve Bayes merupakan metode pembelajaran mesin yang kuat dan sering digunakan dalam klasifikasi serta pemodelan prediktif yang sederhana dan cepat dalam memprediksi kelas dari kumpulan data sampel . Menurut penelitian yang dilakukan oleh [3] pada tahun 2024 tentang Perbandingan Algoritma Klasifikasi Analisis Sentimen Pengguna Aplikasi GetContact Dalam Pencegahan Penipuan Online, hasil nya menunjukkan bahwa Naïve Bayes memiliki akurasi 82,97% sedangkan SVM mencapai 78,00%. Penelitian yang dilakukan oleh [4] pada tahun 2023 tentang Perbandingan Naïve Bayes Dan K-Nearest Neighbor untuk Sentimen Terhadap Ulasan Aplikasi Thread, hasil penelitian ini menunjukkan bahwa Naïve Bayes lebih tinggi akurasinya mencapai 84% sedangkan K-NN hanya mencapai 73%. Penelitian yang dilakukan oleh [5] pada tahun 2022 tentang perbandingan naive bayes dan decision tree menunjukkan bahwa naive bayes akurasinya lebih tinggi yaitu 95.161% sedangkan decision tree akurasinya lebih kecil yaitu hanya 84.677%.

Metode Synthetic Minority Over-sampling Technique(SMOTE) merupakan salah satu algoritma yang digunakan untuk mengatasi imbalanced data. Teknik ini merupakan pendekatan untuk menyeimbangkan data sampel pada kelas yang memiliki ketidakseimbangan yang berlebihan dengan fokus pada kelas minoritas, bertujuan untuk meningkatkan kinerja klasifikasi dataset.

Berdasarkan permasalahan diatas penelitian ini bertujuan untuk menganalisis opini masyarakat tentang liga korupsi di Indonesia, berdasarkan data yang diambil dari salah satu media sosial yaitu youtube. penelitian ini menggunakan metode Naive Bayes yang digunakan untuk mengklasifikasi data seperti teks dan metode SMOTE untuk membuat data menjadi seimbang. Diharapkan hasil dari penelitian ini memberikan gambaran menyeluruh mengenai persepsi masyarakat.

2. Tinjauan Pustaka

2.1. Tindak Pidana Korupsi

Pengertian tindak pidana korupsi menurut Undang-Undang Republik Indonesia No, 31 tahun 1999 pasal 2 ayat (1) dan pasal 3 tentang pemberantasan tindak pidana korupsi adalah sebagai berikut:

1. Pasal 2 ayat (1) : "Setiap orang yang secara melawan hukum melakukan perbuatan memperkaya diri sendiri atau orang lain atau suatu korporasi yang dapat merugikan keuangan negara atau perekonomian negara, dipidana penjara dengan penjara seumur hidup atau pidana penjara paling singkat 4 (empat) tahun dan paling lama 20 (dua puluh) tahun dan denda paling sedikit Rp200.000.000,00 (dua ratus juta rupiah) dan paling banyak Rp1.000.000.000,00 (satu milyar rupiah)"[6].
2. Pasal 3 : "Setiap orang yang dengan tujuan menguntungkan diri sendiri atau orang lain atau suatu korporasi, menyalahgunakan kewenangan, kesempatan atau sarana yang ada padanya karena jabatan atau kedudukan yang dapat merugikan keuangan negara atau perekonomian negara, dipidana dengan pidana penjara seumur hidup atau pidana penjara paling singkat 1 (satu) tahun dan paling lama 20 (dua puluh) tahun dan atau denda paling sedikit Rp50.000.000,00 (lima puluh juta rupiah) dan paling banyak Rp1.000.000.000,00 (satu milyar rupiah)"[6].

2.2. Analisis Sentimen

Analisis sentimen adalah suatu proses ekstraksi opini atau emosi yang terkandung dalam sebuah teks. metode ini digunakan untuk mengidentifikasi sebuah pernyataan apakah positif atau negatif. Dalam penelitian ini analisis sentimen digunakan untuk mengetahui opini publik di salah satu platform media sosial yaitu youtube terhadap kasus liga korupsi Indonesia[7].

2.3. Youtube

Youtube adalah platform media sosial yang menyediakan konten dalam bentuk visual dan audio. Youtube sangat populer dikalangan generasi muda, kemampuannya menyajikan visualisasi yang dinamis dan langsung dapat ditonton. Menurut sianipar (2013), youtube merupakan basis data video paling terkenal di internet dan menjadi situs andalan yang menyediakan beragam informasi dalam bentuk video. Platform ini dirancang khusus untuk pengguna yang ingin mencari dan mengakses berbagai jenis konten video secara langsung[8].

2.4. Algoritma Naive Bayes

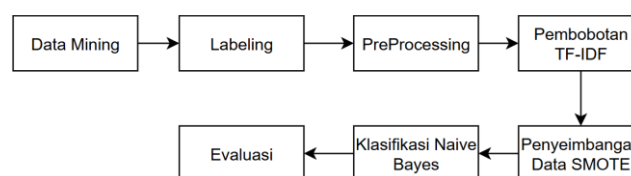
Algoritma Naive Bayes adalah metode yang digunakan dalam teknik klasifikasi data mining. Naive Bayes diterapkan dengan memanfaatkan Teorema Bayes yang diperkenalkan oleh Thomas Bayes, dengan cara memprediksi peluang di masa depan berdasarkan pengalaman yang terjadi sebelumnya. Klasifikasi menggunakan Naive Bayes memungkinkan proses klasifikasi dilakukan secara sederhana namun tetap efektif[9].

2.5 SMOTE (Synthetic Minority Over-Sampling Technique)

SMOTE merupakan metode oversampling yang dirancang untuk mengatasi permasalahan ketidakseimbangan kelas di dalam sebuah dataset. Teknik ini bekerja dengan menghasilkan data sintetik untuk kelas minoritas melalui proses interpolasi antar data. SMOTE dapat menciptakan contoh baru yang mirip namun tidak identik yang dapat meningkatkan keberagaman data minoritas.

3. Metode Penelitian

pada tahap ini membahas seluruh proses yang dilakukan untuk melihat hasil analisis sentimen publik terhadap liga korupsi di Indonesia menggunakan Naive Bayes yang dapat dilihat pada gambar 1.



Gambar 1. Tahapan Penelitian

3.1. Data Mining

Pada tahap ini kami melakukan pengumpulan data melalui komentar di media sosial youtube pada postingan channel METRO TV yang berjudul "LIGA KORUPSI INDONESIA | KONTROVERSI" menggunakan teknik scraping. Data scraping adalah sebuah teknik yang memungkinkan seseorang untuk mengekstrak data dari sebuah situs atau website yang berbeda-beda. Pada tahap ini kami mengambil sebanyak 3,436 komentar, tetapi kami hanya mengambil sebanyak 1,500 komentar untuk di analisis dan untuk menghasilkan dataset yang akan digunakan untuk pre-processing.

3.2. Labeling

Pada tahapan labeling dilakukan dengan menggunakan cara manual labeling. Manual labeling dilakukan dengan cara mengambil 1.500 data komentar yang telah dikumpulkan dari youtube dan diberi label positif dan negatif. Kelas Positif berisi komentar yang mengandung unsur pujian, sedangkan kelas Negatif berisi komentar yang mengandung unsur ketidakpuasan.

3.3. Preprocessing

Preprocessing adalah teknik awal data mining yang bertujuan untuk mengubah data mentah yang diperoleh dari sebuah sumber menjadi informasi yang bersih, terstruktur dan siap digunakan untuk analisis [11]. pada tahap ini terdapat 5 tahapan yaitu sebagai berikut :

1. Case Folding, yaitu proses mengubah semua huruf besar dalam teks menjadi huruf kecil.
2. Cleaning, yaitu proses membersihkan data dengan menghapus yang tidak dibutuhkan, seperti menghapus URL , menghapus hastag, menghapus username, dan menghapus symbol.
3. Tokenizing, yaitu proses yang dilakukan untuk memecahkan kalimat menjadi kata - kata.
4. Stop Word, yaitu proses menghilangkan kata - kata yang banyak muncul dalam teks tetapi tidak memberikan informasi penting.
5. Stemming, yaitu proses mengurangi kata ke bentuk semula.

3.4. Term Frequency-Inverse document frequency

Pada tahap ini melakukan penentuan bobot untuk setiap term pada dokumen. Term Frequency yaitu semakin tinggi frekuensi kemunculan term pada sebuah dokumen maka semakin tinggi nilai bobot untuk term itu sendiri. Dapat dilihat pada persamaan (1), (2) dan (3).

$$tf = 0,5 + 0,5 \times \frac{tf}{max(tf)} \dots (1)$$

$$idf_t = \log \left(\frac{D}{df_t} \right) \dots (2)$$

$$W_{a.t} = tf_{a.t} \times idf_{a.t} \dots (3)$$

Keterangan :

D	= Dokumen ke - d
t	= Term ke-t dari dokumen
w	= Bobot ke-d terhadap term ke-t
tf	= Jumlah kemunculan term i dalam dokumen
idf	= Inversed Document Frequency
df	= Banyak dokumen yang mengandung term i

3.5 Synthetic Minority Oversampling Technique (SMOTE)

Metode smote adalah metode yang digunakan untuk mengatasi data yang tidak seimbang. Metode ini dilakukan dengan menambah jumlah kelas minoritas supaya setara dengan kelas mayoritas. Dapat dilihat pada persamaan (4).

$$X_{synthetic} = Xi + \lambda x (Xj - Xi) \dots (4)$$

Keterangan :

X_i	= Vektor fitur dari titik data minoritas yang dipilih
X_j	= Salah satu tetangga terdekat dari X_i
λ	= Nilai acak antara 0 dan 1 yang menentukan posisi sampel sintetis diantara X_i dan X_j .

3.6 Klasifikasi Naïve Bayes

Naïve Bayes merupakan metode pengklasifikasian probabilistik yang sederhana, yang menghitung probabilitas berdasarkan penjumlahan dari frekuensi dan kombinasi nilai dari dataset. Dapat dilihat pada persamaan (5).

$$P(H|X)^{\wedge} = (P(X|H) \times \frac{P(H)}{P(X)}) \dots (5)$$

Keterangan :

- X = Data class yang belum diketahui
H = Hipotesis data X merupakan suatu class spesifik
 $P(H|X)$ = Probabilitas hipotesis H berdasarkan kondisi X
 $P(H)$ = Probabilitas hipotesis X
 $P(X|H)$ = Probabilitas X berdasarkan kondisi pada hipotesis H
 $P(X)$ = Probabilitas dari X

3.7 Evaluasi

Setelah proses klasifikasi selesai, dilakukan pengujian untuk mengevaluasi kinerja metode yang digunakan menggunakan Confusion Matrix. Confusion matrix adalah matriks untuk perhitungan hasil klasifikasi, dengan menggunakan persamaan berikut :

Tabel 1. Confusion Matrix

Komentar	Prediksi Data	
	Positif	Negatif
Positif	TP	FP
Negatif	FN	TN

4. Hasil Penelitian

4.1. Pengumpulan data

Pengumpulan data diambil dari youtube dengan melakukan scraping menggunakan youtube comment scrapper yang mengambil komentar dari video youtube tertentu dan menyimpannya menjadi dataset. hasil dari pengambilan komentar yang telah menjadi dataset dapat dilihat di Tabel 2.

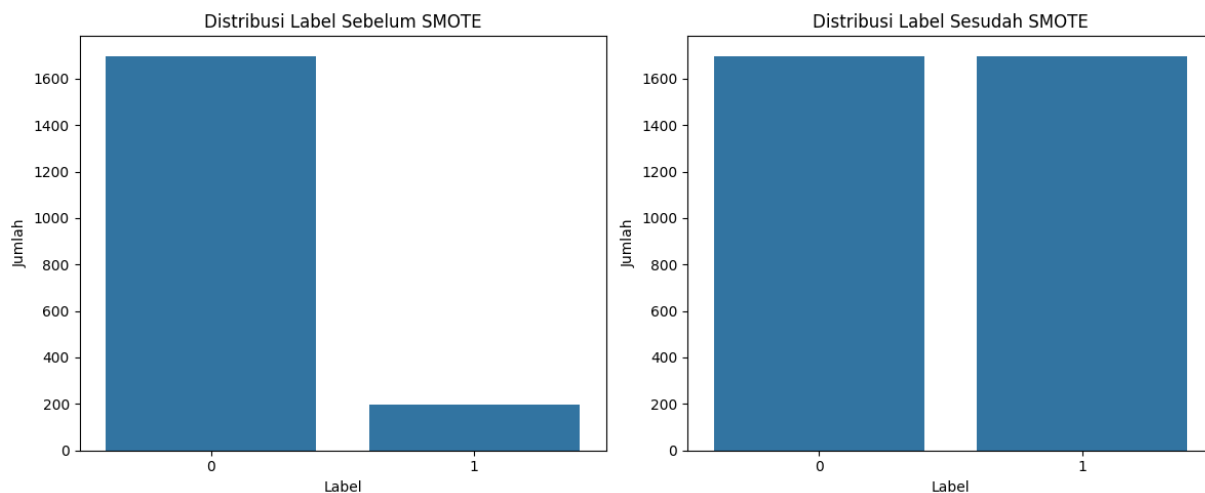
Tabel 2. Hasil Pengambilan Komentar

Comment	Label
Jangan ngomng doang dong 😏, harus pake aksi juga buat memberantas para tikus 😏	Negatif
Bro sedang cooking 🤖	Positif
Ucl korupsi 2024-2025	Negatif
Sistem n aktor sdh pd jerpos n busuk. Aliran dana korupsinya kok ndak pernah dibahas.	Negatif
Hukum mati	Negatif
Suara sama gerak bibir kok nggak sama.	Negatif
Penjara koruptor itu sama gx sih sama penjara pada umumnya?	Negatif
Good	Positif

mantap pak prabowo.rakyat bersama prabowo.	Positif
Menteri BUMN perlu ditangkap dulu...GK ada. BUMN yg untung...Mulyono adili juga.	Negatif

Pada Tabel 2 menampilkan beberapa contoh komentar yang telah diambil dari platform youtube yang dilakukan dengan metode scrapping. Setelah proses pengambilan, setiap komentar di beri label sentimen dengan cara manual yang terbagi menjadi dua kategori yaitu positif dan negatif. Label ini yang akan menjadi target variabel dalam pengembangan model klasifikasi sentimen. Secara keseluruhan, Tabel 2 mempresentasikan tahap awal yang krusial dalam pengolahan data sentimen, yaitu pengumpulan data asli yang relevan dan pemberian label yang tepat sebagai awal sebelum masuk ke tahap selanjutnya.

Setelah proses pengambilan dan pelabelan komentar menjadi positif dan negatif jumlah data yang digunakan adalah sebanyak 1896 setelah proses preprocessing. dari data tersebut, terdapat 1698 sentimen negatif dan 198 sentimen positif. Hal ini menunjukkan bahwa jumlah komentar dengan label negatif (0) jauh lebih banyak dibandingkan komentar dengan label positif (1). Ketidakseimbangan ini berpotensi membuat model lebih memprioritaskan kelas mayoritas (negatif) dan kurang sensitif terhadap kelas minoritas (positif). untuk mengatasi ketidakseimbangan data ini maka dilakukan teknik penyeimbangan menggunakan SMOTE. Berikut perbandingan data sebelum dan sesudah di SMOTE.



Gambar 3. Visualisasi sentimen sebelum dan sesudah di SMOTE.

4.2. Hasil Analisis

Setelah melakukan tahap pengumpulan dan pelabelan data, langkah berikutnya adalah penerapan algoritma Naive Bayes untuk menghasilkan model yang dapat membedakan data training dan data uji. Dari proses training yang sudah dilakukan didapatkan hasil untuk masing-masing kategori 0,1044 untuk positif dan 0,8955 untuk negatif, hal ini menunjukkan bahwa kategori negatif adalah kategori yang paling dominan dalam dataset yang sudah dikumpulkan. Kemudian kategori tersebut, divisualisasikan menggunakan word cloud berdasarkan frekuensi kemunculan kata-kata dalam tiap kategori yang dapat dilihat pada gambar 3.



Gambar 4. Wordcloud

4.3. Evaluasi Model

Metode Naive Bayes dievaluasi menggunakan cross-validation, untuk menguji model Naive Bayes menggunakan SMOTE dan model Naive Bayes Tanpa SMOTE. Berikut adalah perbandingan tabel yang berisi *accuracy*, *precision*, *recall*, *F1-Score* sebelum dan setelah di SMOTE.

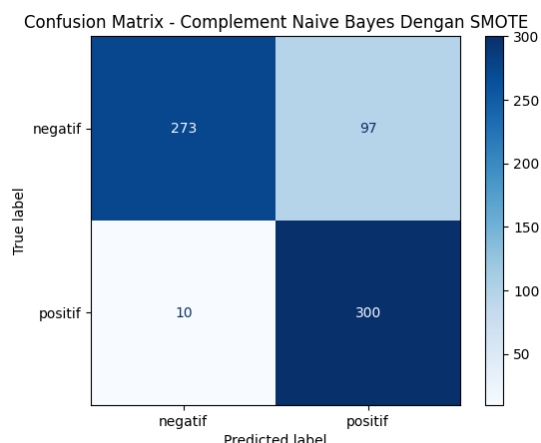
Tabel 3. Accuracy, Precision, Recall, dan F!-score pada Naive Bayes sebelum SMOTE

Accuracy Tanpa Smote : 0.7973684210526316 Classification Report :				
precision		Recall	f1 - score	support
negatif	0.92	0.85	0.88	344
positif	0.17	0.31	0.22	36
accuracy			0.80	380
macro avg	0.55	0.58	0.55	380
weighted avg	0.85	0.80	0.82	380

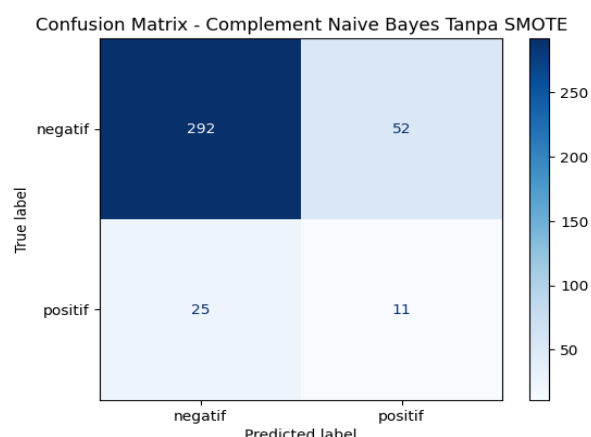
Tabel 4. Accuracy, Precision, Recall, dan F!-score pada Naive Bayes sesudah SMOTE

Accuracy Tanpa Smote : 0.8426470588235294 Classification Report :				
precision		Recall	f1 - score	support
negatif	0.96	0.74	0.84	370
positif	0.76	0.97	0.85	310
accuracy			0.84	680
macro avg	0.86	0.85	0.84	680
weighted avg	0.87	0.84	0.84	680

Berdasarkan hasil yang telah ditunjukkan pada Tabel 1 dan Tabel 2 dapat diperhatikan bahwa akurasi tertinggi berhasil dicapai saat metode Naive Bayes menggunakan SMOTE yaitu 84.2% sedangkan untuk model yang tidak menggunakan SMOTE hanya mencapai akurasi 79.7%. Perbedaan ini menunjukkan bahwa SMOTE berhasil menangani masalah ketidakseimbangan kelas, sehingga meningkatkan kemampuan model dalam mengenali kedua kategori, khususnya kelas minoritas. Selain itu, evaluasi lain seperti precision, recall dan F!-Score juga meningkat saat menggunakan SMOTE dan membuktikan bahwa model SMOTE membuat data menjadi lebih seimbang dan akurat dalam klasifikasi sentimen. Berikut adalah perbandingan confusion matrix menggunakan model Naive Bayes dengan SMOTE dan tanpa SMOTE



Gambar 4. Confusion Matrix Dengan Smote



Gambar 5. Confusion Matrix Tanpa SMOTE

Berdasarkan gambar 4 dan gambar 5 dapat dilihat bahwa model Complement Naive Bayes yang menggunakan SMOTE berhasil mengklasifikasikan 273 data sebagai true negatif dan 300 data sebagai true positif. Sedangkan, model yang tidak menggunakan SMOTE mengklasifikasikan 292 data true negatif dan hanya 11 data sebagai true positif. Dari perbandingan ini terlihat bahwa model dengan SMOTE memiliki keseimbangan yang lebih baik terutama dalam mengenali data positif yang merupakan kelas minoritas. Sebaliknya model tanpa SMOTE cenderung fokus ke kelas mayoritas, yang menyebabkan rendahnya jumlah true positif dan tingginya true negatif. Secara keseluruhan, penggunaan SMOTE terbukti memberikan peningkatan pada model Complement Naive Bayes, dengan nilai lebih seimbang di setiap kelasnya.

5. Kesimpulan

Dari penelitian yang sudah dilakukan menunjukkan bahwa opini publik terhadap liga Korupsi di Indonesia yang di ambil dari platform youtube didominasi oleh sentimen negatif. Dengan menggunakan algoritma Naive Bayes dan SMOTE untuk menyeimbangkan data, klasifikasi mampu mencapai accuracy sebesar 84% dengan nilai precision untuk kelas negatif sebesar 0.96 dan untuk kelas positif sebesar 0.76 serta nilai recall untuk kelas negatif sebesar 0.74 dan kelas positif sebesar 0.97 Hal ini menunjukkan bahwa metode yang digunakan efektif dalam menganalisis teks pada data sosial media. Hasil ini juga menunjukkan keresahan masyarakat terhadap liga korupsi yang ada di Indonesia karena komentar-komentar didominasi oleh negatif. untuk penelitian selanjutnya , disarankan untuk memperluas sumber dan membandingkan menggunakan algoritma yang berbeda.

Daftar Pustaka

- [1] Kementerian Keuangan Republik Indonesia, "Tindak pidana korupsi: Pengertian dan unsur-unsurnya," *Direktorat Jenderal Perbendaharaan*, Feb. 22, 2022. [Online]. Available: <https://djpb.kemenkeu.go.id/kppn/manokwari/id/data-publikasi/berita-terbaru/3026-tindak-pidana-korupsi-pengertian-dan-unsur-unsurnya>.
- [2] Kompas.com, "Daftar megakorupsi dalam klasemen Liga Korupsi Indonesia, terbesar hampir Rp 40 triliun," *Kompas*, Feb. 27, 2025. [Online]. Available: <https://www.kompas.com/tren/read/2025/02/27/161500165/daftar-megakorupsi-dalam-klasemen-liga-korupsi-indonesia-terbesar-hampir-rp?page=all#page2>
- [3] KH. Hermanto, R. Fahlapi, A. Y. Kuntoro, and T. Asra, "Perbandingan algoritma klasifikasi analisis sentimen pengguna aplikasi Getcontact dalam pencegahan penipuan online," *J-INTECH (Journal of Information and Technology)*, vol. 12, no. 1, pp. 158–167, Jun. 2024. [Online]. Available: <https://jurnal.stiki.ac.id/J-INTECH/article/download/1262/773/>
- [4] F. T. Berton, D. E. Ratnawati, and M. A. Rahman, "Perbandingan Naïve Bayes dan K-Nearest Neighbor untuk Analisis Sentimen Terhadap Ulasan Aplikasi Threads," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 1, no. 1, pp. 1–10, Jan. 2017. [Online]. Available: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/download/14109/6301/100161>
- [5] M. A. Ramadhan dan M. I. Wahyudin, "Analisis Sentimen Mengenai Keberhasilan Indonesia di Ajang Thomas Cup 2020 (Studi Kasus Media Sosial Twitter) Menggunakan Metode Naïve Bayes dan Decision Tree," *Jurnal JTik (Jurnal Teknologi Informasi dan Komunikasi)*, vol. 6, no. 4, pp. 505–511, Okt. 2022. [Online]. Available: <https://doi.org/10.35870/jtik.v6i4.560>

- [6] Pemerintah Republik Indonesia, "Undang-Undang Nomor 31 Tahun 1999 tentang Pemberantasan Tindak Pidana Korupsi," *Hukumonline*, 1999. [Online]. Available: <https://www.hukumonline.com/pusatdata/detail/1371/undangundang-nomor-31-tahun-1999/document/>
- [7] H. Fitri and A. Wahyuni, "Analisis pengaruh e-government dalam pencegahan korupsi," *Jurnal Teknik Informatika dan Teknologi Informasi (JITET)*, vol. 9, no. 1, pp. 33–41, 2023. [Online]. Available: <https://journal.eng.unila.ac.id/index.php/jitet/article/view/6399/2533>
- [8] A. H. Rahman, "Pemberantasan korupsi di Indonesia melalui strategi pendidikan antikorupsi," *Rechtsidee Law Journal*, vol. 7, no. 1, pp. 15–22, 2020. [Online]. Available: <https://e-journal.unair.ac.id/RLJ/article/download/8000/8356/55340>
- [9] F. Hidayat, "Peran pendidikan karakter dalam menanggulangi korupsi sejak dini," *Jurnal Pendidikan*, vol. 4, no. 2, pp. 50–55, 2021. [Online]. Available: <https://repository.unmuhjember.ac.id/2730/8/Jurnal.pdf>
- [10] D. Septiani dan I. Isabela, "Analisis Term Frequency Inverse Document Frequency (TF-IDF) dalam Temu Kembali Informasi pada Dokumen Teks," *SINTESIA: Sistem dan Teknologi Informasi Indonesia*, vol. 1, no. 2, pp. 123–130, 2022. [Online]. Available: <https://journal.unj.ac.id/unj/index.php/SINTESIA/article/view/39364>.
- [11] D. Darmanto, N. I. Pradasari, and E. Wahyudi, "Sistem deteksi plagiarisme tugas akhir mahasiswa berbasis Natural Language Processing menggunakan algoritma Jaro-Winkler dan TF-IDF," *Smart Comp: Jurnalnya Orang Pintar Komputer*, vol. 13, no. 1, pp. 201–211, 2024.
- [12] M. P. Pulungan, A. Purnomo, dan A. Kurniasih, "Penerapan SMOTE untuk Mengatasi Imbalance Class dalam Klasifikasi Kepribadian MBTI Menggunakan Naive Bayes Classifier," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, vol. 11, no. 5, pp. 1033–1042, Okt. 2024. [Online]. Available: <https://jtiik.ub.ac.id/index.php/jtiik/article/view/7989>
- [13] S. Rahayu dan A. S. RMS, "Penerapan Metode Naive Bayes Dalam Pemilihan Kualitas Jenis Rumput Taman CV. Rumput Kita Landscape," *Digital Zone: Jurnal Teknologi Informasi dan Komunikasi*, vol. 9, no. 2, pp. 162–171, Nov. 2018. [Online]. Available: https://www.researchgate.net/publication/328931759_Penerapan_Metode_Naive_Bayes_Dalam_Pemilihan_Kualitas_Jenis_Ru
- [14] Caroline, F., Budi, R. G. S., & Rivan, M. E. A. (2024). Analisis sentimen masyarakat terhadap kasus korupsi PT. Timah menggunakan metode Support Vector Machine. *Jurnal Ilmu Komputer dan Informatika (JIKI)*, 4(1), 43–50. [Online]. Available: <https://jiki.jurnal-id.com/index.php/jiki/article/view/141>
- [15] Zulzila, A., Febiola, L. J., & Vionanda, D. (2025). Analysis of public sentiment towards corruption based on tweets using Naive Bayes classifier. *UNP Journal of Statistics and Data Science*, 3(1), 72–78. [Online]. Available: <https://ujdsd.ppj.unp.ac.id/index.php/ujdsd/article/view/345>
- [16] Simanjuntak, D. Y., Prabowo, R., & Sanjaya, A. P. (2023). Analisis sentimen masyarakat terhadap pemerintah pada Twitter menggunakan metode SVM dan Naive Bayes. *Jurnal Teknologi dan Sistem Komputer Terpadu (JITeT)*, 1(2), 139–148. [Online]. Available: <https://journal.eng.unila.ac.id/index.php/jitet/article/view/6399>