Analisis Sentimen Pengguna *Youtube* Terhadap Penyedia Layanan Internet *Starlink* di Indonesia Menggunakan Algoritma *Decision Tree*

1st Haycal Saptahadi Hery Syahputra
Fakultas Rekayasa Industri
Telkom University Purwokerto
Purwokerto, Indonesia
haycalsaptahadi@student.telkomuniversity.ac.id

2nd Sena Wijayanto, S.Pd., M.T.
Fakultas Rekayasa Industri
Telkom University Purwokerto
Purwokerto, Indonesia
senawijayanto@telkomuniversity.ac.id

***Abstrak* — Layanan internet satelit *Starlink* yang hadir di Indonesia menimbulkan berbagai opini di media sosial, khususnya *Youtube*, sehingga penting untuk memahami persepsi masyarakat guna mengevaluasi penerimaan dan tantangan layanan tersebut. Topik ini dingkat karena masalah akses internet di wilayah terpencil Indonesia masih terbatas, sementara pemahaman tentang tanggapan masyarakat terhadap *Starlink* belum terkelola secara sistematis, sehingga menyulitkan penyedia layanan dalam meningkatkan kualitas dan strategi. Penelitian ini menggunakan komentar dari 10 video *Youtube* terkait *Starlink* yang melalui tahap *preprocessing* meliputi pembersihan data, *case folding*, normalisasi, tokenisasi, *stopword removal* dan *stemming*. Data diberi label sentimen positif dan negatif menggunakan kamus *senticnet*, dibobot dengan *TF-IDF*, diseimbangkan dengan teknik *SMOTE*, dan diklasifikasikan menggunakan algoritma *Decision Tree*. Model dievaluasi dengan metrik akurasi, *precision*, *recall*, dan *F1-score*. Hasil menunjukkan mayoritas komentar bersifat positif dengan presentase 59,69% untuk sentimen positif dan 40,31% untuk negatif, dan model *Decision Tree* mencapai akurasi 76,35% menggunakan parameter *default* dan proporsi *split data* 80:20 dan 71,26% setelah *tuning* parameter dan proporsi *split data* 80:20. Penggunaan *tuning* parameter dilakukan untuk mengatasi permasalahan *overfitting* pada data. Penelitian ini diharapkan memberikan wawasan mengenai pandangan masyarakat terhadap penyedia layanan internet *Starlink* di Indonesia.**

Kata kunci*— Analisis Sentimen, *Starlink*, *Decision Tree

I. PENDAHULUAN

Perkembangan teknologi yang pesat telah menjadikan internet sebagai kebutuhan utama bagi individu maupun bisnis. Namun, tantangan geografis dan infrastruktur yang belum merata di Indonesia membuat akses internet di wilayah terpencil masih terbatas[1]. *Starlink*, layanan internet berbasis satelit milik *SpaceX*, hadir sebagai solusi alternatif dengan jangkauan global, termasuk ke daerah-daerah sulit dijangkau[2]. Meskipun menawarkan potensi besar, kehadiran *Starlink* juga menimbulkan pro dan kontra di masyarakat, mulai dari kekhawatiran terhadap dominasi pasar hingga isu privasi data[3].

Youtube sebagai *platform* media sosial yang banyak digunakan di Indonesia[4] menyimpan opini publik dalam bentuk komentar yang dapat dianalisis untuk memahami persepsi masyarakat[5]. Salah satu pendekatan yang digunakan adalah analisis sentimen berbasis teks[6]. Penelitian ini memanfaatkan algoritma *Decision Tree* karena kemampuannya yang tinggi dalam klasifikasi teks dan interpretasi hasil yang mudah[7]. Beberapa penelitian terdahulu menunjukkan bahwa *Decision Tree* memiliki performa yang kompetitif dalam klasifikasi sentimen dibandingkan algoritma lain seperti SVM, KNN, dan *Naïve Bayes*[8][9].

Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap layanan *Starlink* melalui komentar *Youtube*. Dengan pendekatan text mining dan algoritma *Decision Tree*, diharapkan hasil penelitian ini dapat memberikan wawasan bagi penyedia layanan dan pemangku kebijakan dalam menyusun strategi layanan internet berbasis satelit di Indonesia.

II. KAJIAN TEORI

A. *Starlink*

Starlink merupakan sebuah sistem konstelasi satelit yang dikembangkan oleh perusahaan *Space Exploration Technologies Corp.* (*SpaceX*) sebagai penyedia layanan internet global. Teknologi ini dirancang untuk memenuhi kebutuhan akses internet, khususnya di wilayah yang masih sulit terjangkau oleh infrastruktur komunikasi konvensional. Dengan memanfaatkan ribuan satelit yang mengorbit di luar angkasa, *Starlink* mampu memberikan layanan internet yang lebih merata, termasuk ke daerah-daerah terpencil yang sebelumnya kesulitan memperoleh konektivitas [10].

B. Analisis Sentimen

Analisis sentimen merupakan teknik untuk mengolah informasi guna memahami penilaian, opini, sikap, tindakan, serta emosi yang berkaitan dengan suatu topik, layanan, produk, atau permasalahan tertentu[11]. Metode ini secara otomatis mengenali, mengekstrak, dan menganalisis materi berbentuk teks untuk menentukan sentimen yang ada dalam

sudut pandang seseorang. Dengan menggunakan analisis sentimen, kita dapat mengetahui kecenderungan opini seseorang, apakah mengarah pada pandangan positif atau negatif terhadap suatu isu atau objek[12].

C. *Decision Tree*

Pohon keputusan adalah struktur yang mengaplikasikan deretan aturan keputusan untuk memecah koleksi data yang sangat besar menjadi kelompok-kelompok catatan yang lebih kecil[13]. Karena pohon keputusan mudah dipahami dan ditafsirkan oleh orang-orang, maka pohon keputusan merupakan salah satu algoritme kategorisasi yang paling banyak digunakan. Algoritma ini merupakan komponen dari proses klasifikasi data mining. Konsep *Decision Tree* mirip dengan diagram alir dan memiliki bentuk seperti pohon. Setiap simpul internal merepresentasikan sebuah atribut, sementara daunnya menjelaskan kelas atau distribusi kelas dan cabang-cabangnya menampilkan nilai atribut atau hasil pengujian. Beberapa metode, seperti ID3, CART, dan C4.5 sering digunakan untuk membangun pohon keputusan[14].

D. *Youtube*

Youtube merupakan *platform* media sosial populer yang menyediakan berbagai jenis video, mulai dari musik, berita, film, hingga konten buatan para penggunanya. *Youtube* adalah situs *web* yang memungkinkan orang memposting, menonton, dan membagikan video secara gratis[15]. *Platform* ini menyediakan fitur menarik yang meningkatkan antusiasme pengguna, salah satunya adalah fitur interaktif. Melalui fitur ini, penonton dapat memberikan komentar pada video, dan kreator atau pengunggah video memiliki kesempatan untuk merespons komentar tersebut, sehingga menciptakan pengalaman yang lebih interaktif[16].

E. *SMOTE*

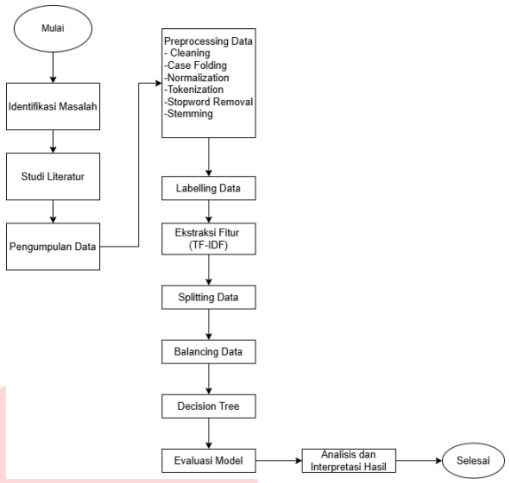
SMOTE (*Synthetic Minority Over-sampling Technique*) merupakan cara yang diterapkan untuk menyeimbangkan data dengan cara mengambil sampel dari kelas yang memiliki jumlah data paling sedikit. Teknik ini bertujuan untuk memastikan bahwa model pembelajaran mesin dapat mengenali pola dari semua kelas secara optimal, termasuk kelas dengan jumlah data yang lebih kecil. Dengan *SMOTE*, model tidak hanya fokus pada kelas dengan data terbanyak tetapi juga mampu memahami pola dari kelas dengan data yang lebih sedikit, sehingga meningkatkan kinerja dalam proses pembelajaran[5].

F. *Confusion matrix*

Confusion matrix merupakan sebuah alat yang dapat digunakan untuk mengevaluasi seberapa baik sebuah classifier dapat mengidentifikasi tuple dari berbagai kelas. Pengukuran yang diterapkan *Confusion matrix* yaitu dengan menghitung *accuracy*, *precision*, *recall*, *f1-score* yang mengacu pada nilai *True Positive* (TP), True Negative (TN), *False Positive* (FP), dan *False Negative* (FN) yang merupakan keluaran dari *Confusion matrix*[17].

III. METODE

Penelitian ini akan dilakukan melalui beberapa tahapan seperti identifikasi masalah, studi literatur, pengumpulan data, *preprocessing*, pelabelan data, ekstrasi fitur *TF-IDF*, *split* data, *balancing data* (*SMOTE*), pemodelan *Decision Tree*, evaluasi model, dan analisis dan Interpretasi hasil.



GAMBAR 1
DIAGRAM ALIR PENELITIAN

Berdasarkan Sistematika Penelitian pada Gambar 1, tahapan penelitian ini dimulai dari identifikasi masalah yang berfokus pada fenomena masuknya layanan internet satelit *Starlink* di Indonesia serta pentingnya memahami respons masyarakat terhadapnya. Selanjutnya dilakukan studi literatur untuk mengumpulkan teori yang relevan, dilanjutkan dengan proses pengumpulan data komentar dari 10 video *Youtube* menggunakan teknik *crawling* melalui *Google Colab*. Data yang diperoleh kemudian melalui tahapan *preprocessing* berupa *cleaning*, *case folding*, normalisasi, tokenisasi, *stopword removal*, dan *stemming*. Selanjutnya, dilakukan pelabelan sentimen menggunakan kamus *senticnet* dan proses ekstraksi fitur dengan metode *TF-IDF* untuk mendapatkan bobot kata. Setelah data berbentuk vektor, dilakukan proses pembagian data (*split*) menjadi data latih dan uji dengan proporsi 80:20 dan 70:30, kemudian diseimbangkan menggunakan metode *SMOTE* untuk mengatasi ketidakseimbangan kelas. Model klasifikasi dibangun menggunakan algoritma *Decision Tree* dengan pengujian parameter *default* dan *tuning* (*max_depth*, *min_samples_split*, dan *min_samples_leaf*) untuk mengurangi *overfitting*. Model kemudian dievaluasi menggunakan metrik akurasi, *precision*, *recall*, dan *f1-score*. Tahap akhir adalah analisis dan interpretasi hasil untuk menilai distribusi sentimen serta efektivitas model dalam mengklasifikasikan opini masyarakat terhadap layanan internet *Starlink*.

IV. HASIL DAN PEMBAHASAN

A. Identifikasi masalah

Penelitian ini dilatarbelakangi dengan fenomena masuknya teknologi baru yaitu internet satelit di Indonesia. Dengan masuknya *Starlink*, muncul sebuah solusi untuk mengatasi persebaran internet di wilayah terpencil di Indonesia. Seiring dengan hal tersebut, respon masyarakat sangat beragam terutama di beberapa *platform* media sosial, salah satunya adalah *Youtube*. Hal tersebut menjadi landasan untuk dilakukan proses analisis sentimen terhadap penyedia

layanan internet *Starlink* dengan menggunakan algoritma *Decision Tree*.

B. Studi literatur

Studi literatur pada penelitian ini dilakukan untuk mengumpulkan teori teori yang relevan terhadap penelitian ini. Beberapa teori tentang objek penelitian, algoritma yang digunakan, sumber data, dan metodologi yang digunakan akan dimasukkan karena relevan terhadap penelitian yang dilakukan

C. Pengumpulan data

Proses pengumpulan data pada penelitian ini dilakukan dengan teknik *crawling* data memanfaatkan bahasa pemrograman *python* dengan bantuan *tools google colab* untuk mendapatkan data secara efisien. Komentar yang digunakan pada penelitian ini bersumber dari 10 video *Youtube* yang menguraikan informasi mengenai layanan internet *Starlink* di Indonesia. Video *Youtube* yang digunakan berasal dari *channel channel* berita populer seperti CNN Indonesia, CNBC Indonesia, dan juga Metro TV. Video video tersebut dipilih karena memiliki jumlah data yang sesuai dan tingkat relevansi yang tinggi terhadap topik penelitian. Data yang didapatkan setelah proses crawling sebanyak 3636 data yang mencakup informasi mengenai tanggal komentar, username pengguna, komentar, dan jumlah like. Berikut contoh data awal dapat diperhatikan pada Tabel 1.

TABEL 1
CONTOH DATA AWAL

publishedAt	***AuthorDisplayName***	***textDisplay***	***likeCount***
2024-09-06T03:23:19Z	@Liza-i1m9w	Mau di jual Negara ini kahh???ðŸ˜¢ðŸ˜¢ðŸ˜¢	0

D. *Preprocessing* data

Fase *preprocessing* dilakukan serangkaian tahap untuk mempersiapkan data agar menjadi format yang sesuai untuk dilakukan langkah berikutnya, tahap tahap yang dilakukan pada fase *preprocessing* dokumen antara lain sebagai berikut.

1) *Cleaning*

Pada tahap ini, dataset awal akan dilakukan proses pembersihan seperti menghapus tanda baca, angka, spasi ganda dan karakter khusus. Tujuan dari proses ini untuk mengurangi kemungkinan kesalahan dalam analisis sentimen sehingga mendapatkan hasil nilai akurasi yang optimal. Contoh fase *cleaning* data dapat dilihat pada Tabel 2

TABEL 2
CONTOH FASE *CLEANING DATA*

Sebelum *Cleaning*	**Setelah *Cleaning***
Mati provider indonesia.. indonesia negara konsumtif.. ngga ada pikiran buat berkembanag.. yg penting pejabat cuan	Mati provider Indonesia indonesia negara konsumtif ngga ada pikiran buat berkembanag yg penting pejabat cuan

2) *Case Folding*

Pada tahap ini adalah proses mengubah huruf kapital menjadi huruf kecil agar keseluruhan data yang akan diolah memiliki bentuk yang sama. Contoh fase *case folding* dapat dilihat pada Tabel 3.

TABEL 3
CONTOH FASE *CASE FOLDING*

Sebelum *Case folding*	**Setelah *Case folding***
Mati provider Indonesia indonesia negara konsumtif ngga ada pikiran buat berkembanag yg penting pejabat cuan	mati provider indonesia indonesia negara konsumtif ngga ada pikiran buat berkembanag yg penting pejabat cuan

3) Normalisasi

Tahap selanjutnya yaitu normalisasi yang bertujuan untuk mengubah kata-kata dalam dataset masih bersifat tidak baku agar menjadi kata baku. Contoh fase normalisasi dapat dilihat pada Tabel 4.

TABEL 4
CONTOH FASE NORMALISASI

Sebelum Normalisasi	**Setelah Normalisasi**
mati provider indonesia indonesia negara konsumtif ngga ada pikiran buat berkembanag yg penting pejabat cuan	mati provider indonesia indonesia negara konsumtif tidak ada pikiran untuk berkembang yang penting pejabat untung

4) Tokenisasi

Pada tahap ini, kalimat yang terdapat di dataset akan dipisahkan menjadi per kata dengan tujuan mempermudah proses analisis dan pemodelan. Contoh fase tokenisasi dapat dilihat pada Tabel 5.

TABEL 5
CONTOH FASE TOKENISASI

Sebelum Tokenisasi	**Setelah Tokenisasi**
mati provider indonesia indonesia negara konsumtif tidak ada pikiran untuk berkembang yang penting pejabat untung	‘mati’, ‘provider’. ‘indonesia’. ‘indonesia’, ‘negara’, ‘konsumtif’, ‘tidak’, ‘ada’, ‘pikiran’, ‘untuk’, ‘berkembang’, ‘yang’, ‘penting’, ‘pejabat’, ‘untung’

5) *Stopword removal*

Tahapan *stopword* bertujuan menghilangkan kata-kata dari proses analisis dan pemodelan yang tidak memiliki arti signifikan. Contoh fase *stopword removal* dapat dilihat pada Tabel 6.

TABEL 6
CONTOH FASE *STOPWORD REMOVAL*

Sebelum *Stopword Removal*	**Setelah *Stopword Removal***
‘mati’, ‘provider’. ‘indonesia’. ‘indonesia’, ‘negara’, ‘konsumtif’, ‘tidak’, ‘ada’, ‘pikiran’, ‘untuk’, ‘berkembang’, ‘yang’, ‘penting’, ‘pejabat’, ‘untung’	‘mati’, ‘provider’. ‘indonesia’. ‘indonesia’, ‘negara’, ‘konsumtif’, ‘tidak’, ‘ada’, ‘pikiran’, ‘berkembang’, ‘penting’, ‘pejabat’, ‘untung’

6) *Stemming*

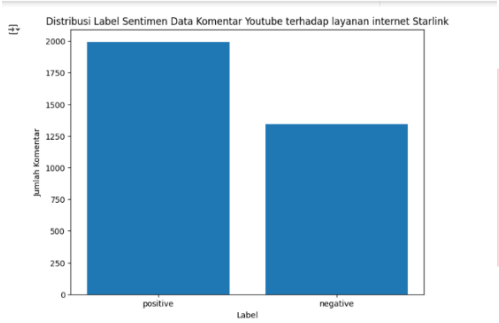
Tahap *stemming* ini bertujuan untuk mengubah kata menjadi bentuk dasar dengan tujuan menyederhanakan dan menyatukan teks yang memiliki arti serupa. Contoh fase *stemming* dapat dilihat pada Tabel 7.

TABEL 7
CONTOH FASE *STEMMING*

Sebelum *Stemming*	**Setelah *Stemming***
‘mati’, ‘provider’. ‘indonesia’. ‘indonesia’, ‘negara’, ‘konsumtif’, ‘tidak’, ‘ada’, ‘pikiran’, ‘berkembang’, ‘penting’, ‘pejabat’, ‘untung’	‘mati’, ‘provider’, ‘indonesia’, ‘indonesia’, ‘negara’, ‘konsumsi’, ‘tidak’, ‘ada’, ‘pikir’, ‘kembang’, ‘penting’, ‘jabat’, ‘untung’

E. *Labelling Data*

Proses pelabelan data pada penelitian ini dilakukan dengan menggunakan kamus *senticnet*, di mana setiap kata dalam komentar diberi nilai (angka) *polarity* berdasarkan kamus tersebut. Komentar yang memiliki *polarity score* > 0 akan diklasifikasikan sebagai komentar positif, sedangkan komentar dengan *polarity score* ≤ 0 akan diklasifikasikan sebagai komentar negatif. *Senticnet* digunakan karena menyediakan skor polaritas secara langsung dalam bentuk angka positif dan negatif, sehingga cocok digunakan untuk klasifikasi dua kelas sentimen, yaitu positif dan negatif, tanpa memerlukan proses konversi atau kategorisasi tambahan. Hasil distribusi label dapat dilihat pada Gambar 2.



GAMBAR 2
DISTRIBUSI LABEL DATA

Gambar 2 merupakan visualisasi hasil pelabelan sentimen menggunakan pendekatan leksikon, data komentar *Youtube* yang dianalisis memperlihatkan bahwa sebagian besar pengguna menyampaikan reaksi yang bersifat positif terhadap penyedia layanan *Starlink* di Indonesia. Dari total keseluruhan data, sebanyak 1.993 komentar atau sekitar 59,69% diklasifikasikan sebagai sentimen positif. Sementara itu, sebanyak 1.346 komentar atau sekitar 40,31% dikategorikan sebagai sentimen negatif.

F. Ekstraksi Fitur (*TF-IDF*)

Penelitian ini menerapkan teknik *Term Frequency Inverse-Document Frequency* (*TF-IDF*) untuk menaksir bobot dari setiap kata dalam dataset. Proses ini bertujuan untuk menentukan seberapa penting sebuah kata dalam data, berdasarkan seberapa sering kata tersebut muncul. Semakin sering sebuah kata muncul, semakin besar bobot atau kontribusinya. Selain itu, *TF-IDF* juga efektif dalam membantu meningkatkan akurasi dan kemampuan model untuk mengenali data dengan lebih baik. *TF-IDF* dipilih karena kemampuannya mengubah data menjadi bentuk numerik dan mampu memfilter kata-kata umum yang tidak terlalu bermakna dalam analisis sentimen, serta memberikan penekanan lebih pada kata kata unik yang dapat membedakan sentimen positif dan negatif secara lebih akurat.

G. *Splitting data*

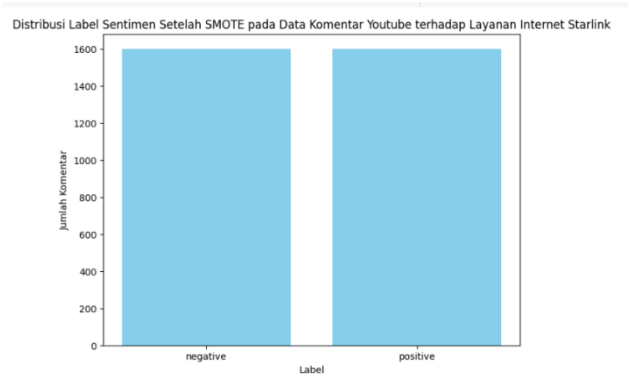
Untuk memisahkan data menjadi data latih dan data uji, prosedur pembagian data diterapkan dalam penelitian ini. Model mempelajari pola sentimen menggunakan data latih, dan kinerjanya terhadap data baru yang belum pernah dilihat sebelumnya dievaluasi menggunakan data uji. Pembagian data dilakukan dengan rasio 80% untuk data latih dan 20% untuk data uji. menghasilkan akurasi yang baik, sebagai perbandingan akan dilakukan pengujian menggunakan proporsi 70% data latih dan 30% data uji. Berikut distribusi data latih dan data uji dapat dilihat pada Tabel 8.

TABEL 8
HASIL *SPLITTING DATA*

Rasio	Data latih	Data Uji
80: 20	2671	668
70:30	2337	1002

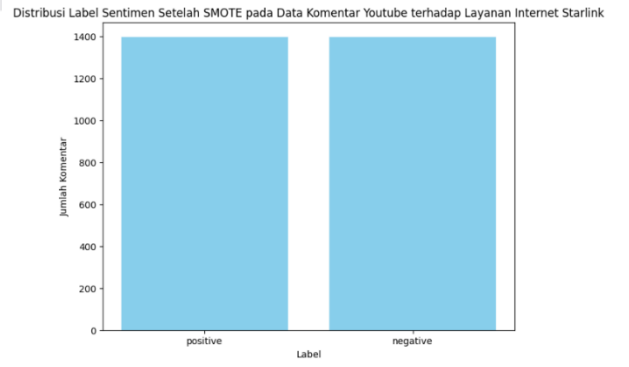
H. *Balancing data*

Setelah Masalah ketidakseimbangan kelas pada data sentimen diatasi melalui prosedur penyeimbangan data yang diterapkan pada penelitian ini. Ketidakseimbangan terjadi saat satu kelas mempunyai jumlah data yang jauh lebih banyak dibandingkan kelas lainnya, yang dapat menyebabkan model cenderung bias terhadap kelas mayoritas. Untuk mengatasi masalah ini, digunakan teknik *SMOTE* (*Synthetic Minority Over-sampling Technique*). *SMOTE* bekerja dengan membuat sampel sintetis untuk kelas minoritas menggunakan data yang sudah tersedia, sehingga menghasilkan jumlah sampel yang seimbang untuk setiap kelas. Dengan data yang seimbang, model dapat belajar mengidentifikasi pola sentimen dengan lebih baik dan memberikan hasil yang lebih optimal. Hasil distribusi data setelah *SMOTE* untuk kedua rasio split data dapat dilihat pada Gambar 3 dan Gambar 4.



GAMBAR 3
DISTRIBUSI HASIL *SMOTE* RASIO 80:20

Gambar 3 merupakan distribusi data setelah dilakukan proses *balancing* menggunakan teknik *SMOTE* pada rasio *splitting* data 80:20 untuk data latih dan data uji, jumlah data pada masing-masing label sentimen menjadi seimbang. Label "*positive*" dan "*negative*" masing-masing memiliki 1.600 data, sehingga total data latih setelah *balancing* berjumlah 3.200. Ini menunjukkan bahwa *SMOTE* berhasil menambahkan data sintetis pada kelas yang sebelumnya lebih sedikit yaitu kelas "*negative*".



GAMBAR 4
DISTRIBUSI HASIL *SMOTE* RASIO 70:30

Gambar 4 merupakan distribusi data setelah dilakukan proses *balancing* menggunakan teknik *SMOTE* pada rasio *splitting* data 70:30 untuk data latih dan data uji, jumlah data pada masing-masing label sentimen menjadi seimbang. Label "*positive*" dan "*negative*" masing-masing memiliki 1397 data, sehingga total data latih setelah *balancing* berjumlah 2794. Ini menunjukkan bahwa *SMOTE* berhasil menambahkan data sintetis pada kelas yang sebelumnya lebih sedikit yaitu kelas "*negativ*e".

I. *Decision Tree*

Pada tahap ini, algoritma *Decision Tree* dengan tipe *CART* (*Classification and Regression Tree*) diterapkan pada data pelatihan menggunakan *library scikitlearn*. Model membangun pohon keputusan berdasarkan fitur-fitur yang dihasilkan dari proses vektorisasi teks, seperti kata-kata atau frasa yang dianggap paling berperan dalam membedakan sentimen. Pada penelitian ini akan dilakukan pemodelan menggunakan parameter *default* bawaan dari *library scikit-learn* dan menggunakan *tuning* parameter yang telah disesuaikan untuk mengatasi masalah *overfitting* pada data. Beberapa parameter yang ditetapkan antara lain seperti Parameter *max_depth=30* untuk membatasi kedalaman maksimum dari pohon keputusan, sehingga model tidak terlalu dalam dan tetap general, parameter *min_samples_split=8* yang berarti sebuah node hanya akan dipecah (split) jika memiliki minimal 8 sampel data, dan *min_samples_leaf=10* agar setiap daun (node akhir) harus memiliki minimal 10.

J. Evaluasi Model

Proses evaluasi model dalam penelitian ini dilakukan untuk mengetahui seberapa baik model *Decision Tree* mampu memprediksi sentimen dari komentar *Youtube* tentang layanan internet *Starlink*. Untuk mengevaluasi, digunakan data uji yang sebelumnya dipisahkan dari data latih. Model dinilai dalam beberapa metrik, yaitu akurasi untuk melihat persentase prediksi yang benar, *precision* untuk mengukur ketepatan prediksi pada sentimen tertentu, *recall* untuk melihat seberapa banyak data yang berhasil diprediksi dengan benar, dan *f1-score* untuk menunjukkan keseimbangan antara *precision* dan *recall*.

1). Evaluasi pemodelan parameter *default*

Pada proses pemodelan awal dilakukan pemodelan algoritma *Decision Tree* menggunakan parameter *default* dari *library scikit-learn*, setelah dilakukan evaluasi menggunakan beberapa metriks evaluasi, didapatkan hasil seperti dapat dilihat pada Tabel 9 dan Tabel 10.

TABEL 9
HASIL EVALUASI PADA DATA *TRAINING*

Rasio	Akurasi	*Precision*	*Recall*	*F1-score*
80: 20	100%	100%	100%	100%
70:30	100%	100%	100%	100%

Tabel 9 merupakan hasil evaluasi pada data *training* dalam proses pemodelan dengan menggunakan parameter *default* dan proporsi *split* data 80:20 dan 70:30, hasil evaluasi menunjukkan bahwa model *Decision Tree* memiliki performa yang sangat tinggi pada data *training*, dengan akurasi sempurna sebesar 100%. Ini berarti model mampu mengklasifikasikan semua data *training* dengan benar tanpa kesalahan. Baik *precision*, *recall*, maupun *f1-score* juga menunjukkan nilai 100%, yang menandakan bahwa model sangat cocok untuk data *training*.

TABEL 10
HASIL EVALUASI PADA DATA *TESTING*

Rasio	Akurasi	*Precision*	*Recall*	*F1-score*
80: 20	76,35%	76,32%	76,35%	76,33%
70:30	73,65%	73,98%	73,65%	73,77%

Tabel 10 merupakan hasil evaluasi pada data *testing* dalam proses pemodelan dengan menggunakan parameter *default*, performa menunjukkan penurunan yang signifikan pada kedua proporsi data. Pada proporsi *split* 80:20, akurasi tercatat sebesar 76,35%, dengan nilai *precision*, *recall*, dan *f1-score* masing-masing sebesar 76,32, ,76,35, dan 73,77. Sementara itu, pada proporsi split 70:30, akurasi sedikit menurun menjadi 73,65%, dengan nilai *precision*, *recall*, dan *f1-score* sebesar 73,98%, 73,65%, dan 73,77%. Selisih akurasi yang cukup besar antara data training dan testing, khususnya pada proporsi 70:30 yang mencapai 26,35% mengindikasikan adanya gejala *overfitting*.

2). Evaluasi pemodelan *tuning* parameter

Pada proses pemodelan selanjutnya dilakukan pemodelan algoritma *Decision Tree* menggunakan parameter *tuning*, beberapa parameter yang digunakan seperti parameter *max_depth=30*, parameter *min_samples_split=8*, dan parameter *min_samples_leaf=10*. Penggunaan parameter dilakukan untuk mengatasi permasalahan *overfitting* yang terjadi pada pemodelan dengan parameter *default*. Setelah dilakukan evaluasi menggunakan beberapa metriks evaluasi, didapatkan hasil seperti dapat dilihat pada Tabel 11 dan Tabel 12.

TABEL 11
HASIL EVALUASI PADA DATA *TRAINING*

Rasio	Akurasi	*Precision*	*Recall*	*F1-score*
80: 20	80,53%	81,05%	80,53%	80,45%
70:30	79,71%	80,90%	79,71%	79,51%

Tabel 11 merupakan hasil pengujian pemodelan pada data *training* menggunakan *tuning* parameter menunjukkan bahwa model *Decision Tree* memberikan performa yang cukup baik dan stabil pada kedua proporsi data. Pada proporsi *split* 70:30, model menghasilkan akurasi sebesar 79,71%, dengan nilai *precision* 80,90%, *recall* 79,71%, dan *f1-score* 79,51%. Sementara itu, pada proporsi *split* 80:20, performa sedikit lebih tinggi dengan akurasi sebesar 80,53%, *precision* 81,05%, *recall* 80,53%, dan *f1-score* 80,45%. Hasil ini menunjukkan bahwa setelah dilakukan *tuning* parameter, model mampu mengenali pola pada data *training* dengan lebih seimbang dan tidak terlalu tinggi seperti pada penggunaan parameter *default*. Performa yang konsisten pada kedua proporsi data juga mengindikasikan bahwa model memiliki generalisasi yang lebih baik terhadap data *training*.

TABEL 12
HASIL EVALUASI PADA DATA *TESTING*

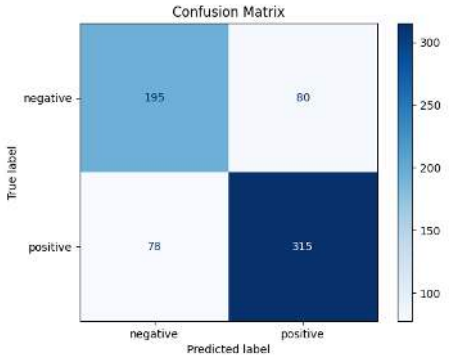
Rasio	Akurasi	*Precision*	*Recall*	*F1-score*
80: 20	71,26%	73,43%	71,26%	71,47%
70:30	70,56%	73,51%	70,56%	70,78%

Tabel 12 merupakan hasil pengujian pemodelan pada data *testing* menggunakan *tuning* parameter menunjukkan performa yang cukup stabil pada kedua proporsi pembagian data. Pada rasio 70:30, model menghasilkan akurasi sebesar 70,56%, *precision* 73,51%, *recall* 70,56%, dan *f1-score* 70,78%. Sementara pada rasio 80:20, akurasi sedikit lebih

tinggi yaitu 71,26%, dengan *precision* 73,43%, *recall* 71,26%, dan *f1-score* 71,47%. Performa yang konsisten ini menunjukkan bahwa *tuning* parameter berhasil mengurangi *overfitting* yang terjadi pada parameter *default*, sehingga model dapat bekerja lebih baik dalam mengenali pola data yang belum pernah dilihat sebelumnya.

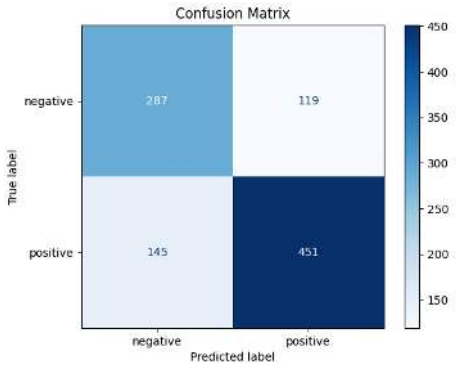
3). *Confusion matrix*

Setelah dilakukan evaluasi model menggunakan beberapa metriks evaluasi, akurasi model juga akan divisualisasikan melalui *confusion matrix* untuk mengetahui seberapa banyak prediksi benar dan salah yang didapatkan dari hasil pemodelan. Hasil *confusion matrix* dapat dilihat pada Gambar 5 hingga Gambar 8.



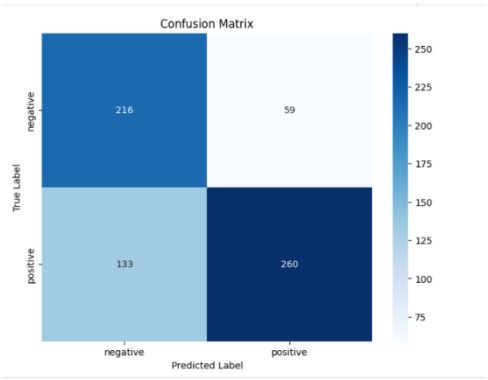
GAMBAR 5
CONFUSION MATRIX PARAMETER *DEFAULT* RASIO 80:20

Gambar 5 merupakan hasil *confusion matrix* dari pemodelan menggunakan parameter *default* dengan proporsi data 80:20. Model berhasil mengklasifikasikan 315 data positif (*true positive*) dan 195 data negatif (*true negative*). Namun, masih terdapat 80 data negatif yang diklasifikasikan sebagai positif (*false positive*) dan 78 data positif diklasifikasikan sebagai negatif (*false negative*).



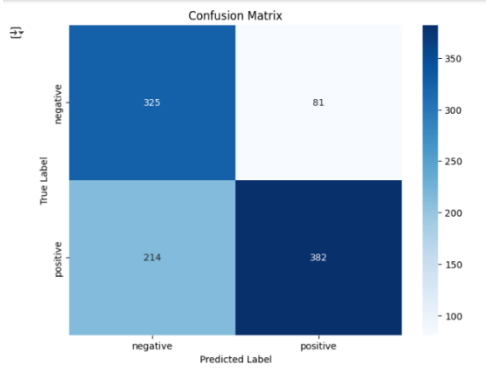
GAMBAR 6
CONFUSION MATRIX PARAMETER *DEFAULT* RASIO 70:30

Gambar 6 merupakan hasil *confusion matrix* dari pemodelan menggunakan parameter *default* dengan proporsi data 70:30. Model berhasil mengidentifikasi 451 komentar positif secara akurat (*true positive*) dan 287 komentar negatif dengan tepat (*true negative*). Namun, terdapat 119 komentar negatif yang salah diklasifikasikan sebagai positif (*false positive*) dan 145 komentar positif terklasifikasi sebagai negatif (*false negative*).



GAMBAR 7
CONFUSION MATRIX PARAMETER *TUNING* RASIO 80:20

Gambar 7 menunjukkan hasil *confusion matrix* dari pemodelan dengan *tuning* parameter dan proporsi data 80:20. Model mampu mengenali 260 komentar positif dengan benar (*true positive*) serta 216 komentar negatif secara tepat (*true negative*), Namun, masih ditemukan 59 komentar negatif yang keliru diprediksi sebagai positif (*false positive*) dan terdapat 133 komentar positif yang teridentifikasi sebagai negatif (*false negative*).

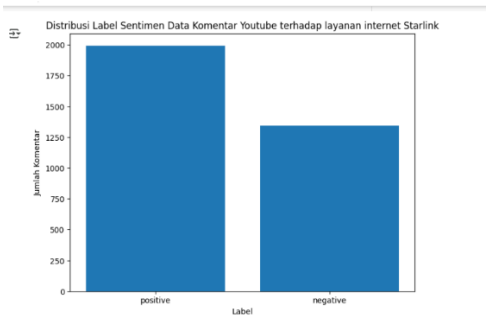


GAMBAR 8
CONFUSION MATRIX PARAMETER *TUNING* RASIO 70:30

Gambar 8 menampilkan hasil *confusion matrix* dari pemodelan dengan *tuning* parameter dan proporsi data 70:30. Sebanyak 382 komentar positif berhasil dikenali dengan benar (*true positive*), dan 325 komentar negatif diklasifikasikan dengan tepat (*true negative*), Namun, terdapat 81 komentar negatif yang salah terbaca sebagai positif (*false positive*), dan 214 komentar positif terprediksi sebagai negatif (*false negative*).

K. Analisis Interpretasi Hasil

Proses analisis dan interpretasi hasil pada penelitian ini dilakukan setelah model *Decision Tree* selesai dievaluasi. Hasil prediksi sentimen dari data uji dianalisis untuk melihat distribusi sentimen positif dan negatif terhadap penyedia layanan internet *Starlink* di Indonesia. Interpretasi hasil melibatkan pengamatan terhadap pola sentimen masyarakat, seperti apakah mayoritas komentar menunjukkan dukungan atau kritik terhadap layanan internet *Starlink*. Berdasarkan hasil, dapat disimpulkan bagaimana pandangan masyarakat terhadap *Starlink*, dan hasil ini dapat menjadi masukan untuk perbaikan atau pengembangan layanan di masa mendatang.



GAMBAR 9
DISTRIBUSI HASIL SENTIMEN

Gambar 9 menunjukkan berdasarkan data komentar *Youtube* yang dianalisis memperlihatkan bahwa sebagian besar pengguna menyampaikan reaksi yang bersifat positif terhadap penyedia layanan *Starlink* di Indonesia. Dari total keseluruhan data, sebanyak 1.993 komentar atau sekitar 59,69% diklasifikasikan sebagai sentimen positif. Sementara itu, sebanyak 1.346 komentar atau sekitar 40,31% dikategorikan sebagai sentimen negatif. Persebaran ini menunjukkan bahwa opini masyarakat yang terekam dalam komentar *Youtube* cenderung lebih mendukung atau memiliki pandangan yang baik terhadap kehadiran layanan *Starlink*, meskipun masih terdapat sejumlah komentar yang bersifat kritis atau negatif.



GAMBAR 10
WORDCLOUD SENTIMEN POSITIF

Pengguna yang memberikan sentimen positif memiliki beberapa alasan seperti dilihat dari *wordcloud* sentimen positif pada Gambar 10 dimana terdapat kata kata yang memiliki konotasi positif seperti “cepat”, “bisa”, dan “bagus”. Beberapa topik yang dibahas pada sentimen positif diantaranya *Starlink* dianggap dapat memberikan solusi terhadap sulitnya persebaran akses internet di Indonesia, karena cara kerja internet satelit yang tidak perlu menggunakan banyak kabel dalam pengaplikasiannya. Penguatan konektivitas pada Ibukota baru Republik Indonesia yaitu Ibukota Nusantara juga menjadi dasar penerimaaan yang baik terhadap penyedia layanan internet *Starlink*, selain itu alasan lain seperti dapat mendukung sarana kesehatan dan pendidikan hingga menjadi pendorong kompetisi yang sehat dan inovatif terhadap penyedia layanan internet lokal juga menjadi kesempatan untuk membuka lapangan pekerjaan bagi masyarakat indonesia. Alasan alasan tersebut menjadi dasar penerimaan yang baik terhadap penyedia layanan internet *Starlink*.



GAMBAR 11
WORDCLOUD SENTIMEN NEGATIF

Di sisi lain, terdapat juga sentimen negatif yang muncul terhadap penyedia layanan internet *Starlink* di Indonesia dilihat dari *wordcloud* sentimen negatif pada Gambar 11 terdapat kata kata yang memiliki konotasi negatif seperti “mahal”, “korupsi”, “monopoli”, dan “saing” beberapa topik yang dibahas dalam komentar diantaranya kekhawatiran masyarakat terhadap terkait kedaulatan data, potensi penyalahgunaan informasi, hingga anggapan bahwa *Starlink* tidak berinvestasi secara langsung di dalam negeri, melainkan hanya menjadikan indonesia sebagai pasar. Beberapa komentar juga menyuarakan kekhawatiran tentang harga layanan yang dianggap mahal, serta kemungkinan melemahnya daya saing provider lokal. Bahkan muncul narasi seperti isu keamanan nasional hingga anggapan bahwa *Starlink* digunakan sebagai alat intai asing

V. KESIMPULAN

Setelah dilakukan penelitian, sentimen berhasil diklasifikasikan ke dalam dua kategori utama positif dan negatif melalui tahapan *preprocessing*, pelabelan menggunakan kamus *senticnet*, pembobotan *TF-IDF*, *balancing* dengan *SMOTE*, dan klasifikasi menggunakan algoritma *Decision Tree*. Hasil penelitian menunjukkan mayoritas komentar masyarakat Indonesia di *Youtube* terhadap penyedia layanan *Starlink* menunjukkan sentimen positif sebesar 59,69%, yang mencerminkan harapan terhadap solusi internet di wilayah terpencil, dukungan terhadap pembangunan IKN, dan peningkatan kualitas layanan internet nasional. Sentimen negatif sebesar 40,31% umumnya berkaitan dengan isu nasionalisme, harga layanan, dan kekhawatiran terhadap keamanan data. Evaluasi menunjukkan bahwa penggunaan parameter *default* dengan proporsi *split* data 80:20 menghasilkan akurasi 100% pada data *training* dan 76,35% pada data *testing*, sedangkan pada proporsi 70:30 menghasilkan akurasi 100% pada data *training* dan 73,65% pada data *testing*. Performa yang sangat tinggi pada data *training* ini mengindikasikan adanya *overfitting*. Setelah dilakukan *tuning* parameter, akurasi menurun menjadi 80,53% pada data *training* dan 71,26% pada data *testing* untuk proporsi 80:20, serta 79,71% pada data *training* dan 70,56% pada data *testing* untuk proporsi 70:30.

REFERENSI

[1] Hukunala, S. V. “*Starlink On Competition Of Internet Providers In Indonesia: A Business Law Review*”. Authentica, 7(1), 1–9. 2024.

[2] Beatrix, S. A., Mikhael, S., Sumarauw, M. Y., & Andreano, W. "Ancaman *Starlink* Terhadap Provider Lokal di Indonesia". Jurnal Ilmiah Sain dan Teknologi (Vol. 3, Issue 1). 2024.

[3] Nudan, P. W., Widodo, P., & Affifudin, M. "*Navigating the Starlink Era of Personal Data Protection in Indonesia*". *Formosa Journal of Science and Technology*, 3(7), 1447–1458. 2024.

[4] Mohammad, W., Saifurrahman, H., Rausyan Fiqrussalam, J., Umar Sulthoni, A., & Fansyiari Austi, D. "Pengaruh Jumlah Video Terhadap Jumlah *Subscribers* dengan *Concurrent Viewers* Sebagai Variabel Intervening Pada *Virtual Youtuber* Indonesia". Triwikrama: Jurnal Multidisiplin Ilmu Sosial. 2023.

[5] Ramadhani, B., & Suryono, R. R. "Komparasi Algoritma *Naïve Bayes* dan *Logistic Regression* Untuk Analisis Sentimen *Metaverse*". Jurnal Media Informatika Budidarma, 8(2), 714. 2024.

[6] Nanda Aulia, Z., Kuncoro Jati, G., & Santoso, I. "Analisis Sentimen Tanggapan Public Mengenai E-Tilang Melalui Media Sosial *Youtube* Menggunakan Algoritma *Naive Bayes*". IKRA-ITH Informatika: Jurnal Komputer Dan Informatika. 2023.

[7] Octa Nuryawan, A. T., Hasbullah, M., Rizal, M., Rajab, M. F., & Agustina, N. "Algoritma *Decision Tree* Untuk Analisis Sentimen Public Terhadap *Marketplace* di Indonesia". Naratif : Jurnal Nasional Riset, Aplikasi Dan Teknik Informatika, 5(1), 18–25. 2023.

[8] Sakinah, D., Azkiah, A., & Hikmah, F. N. "Perbandingan Algoritma *SVM* dan *Decision Tree* Dalam Klasifikasi Kepuasan Pengguna Aplikasi Migo E-Bike di *Playstore". Technology and Science (BITS)*, 6(1). 2024.

[9] Puspita, R., & Widodo, A. "Perbandingan Metode KNN, Decision Tree, dan Naïve Bayes Terhadap Analisis Sentimen Pengguna Layanan BPJS". Jurnal Informatika Universitas Pamulang, 5(4), 646. 2021

[10] Khalil Gibran, M., Rifki, M. I., Hasugian, A. H., Al, A. T., Siahaan, A., Sahputra, A., Ong, & Malahayati, P. P. "*Sentiment Analysis of Platform X Users on Starlink Using Naive Bayes"*. Instal : Jurnal Komputer. 2024.

[11] Maulana, R., Voutama, A., & Ridwan, T. "Analisis Sentimen Ulasan Aplikasi *Mypertamina* Pada *Google"*. Jurnal Teknologi Terpadu, Vol. 9(No. 1 2023), 42–48. 2023.

[12] Gifari, O. I., Adha, M., Rifky Hendrawan, I., Freddy, F., & Durrand, S. "Analisis Sentimen *Review* Film Menggunakan *TF-IDF* dan *Support Vector Machine*". JIFOTECH (JOURNAL OF INFORMATION TECHNOLOGY, 2(1). 2022.

[13] MZ, Y., Bororing, J. E., & Fuadiah, N. "Penerapan Metode *K-Nearest Neighbor* dan *Decision Tree* Untuk Analisis Sentimen (Studi Kasus Mario Dandi)". *Indonesian Journal Of Information Technology*. 2023.

[14] Septhya, D., Rahayu, K., Rabbani, S., Fitria, V., Irawan, Y., & Hayami, R. "Implementasi Algoritma *Decision Tree* dan *Support Vector Machine* untuk Klasifikasi Penyakit Kanker Paru". *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 3, 15–19. 2023.

[15] Mahardhika, A. A. G., D. I., Gede, A. A., & Geriya, M. "*Violation And Copyright Protection Policy On Youtube*". *Jurnal Living Law*, 13(2). 2021.

[16] Prayoga Priansyah, R., Satriani, I. "Pengaruh Terpaan Tayangan *Youtube* Dehakims Terhadap Sikap Orangtua Mengenai Anak Menonton Satwa". Jurnal Penelitian Sosial Ilmu Komunikasi, 6(1), 19–28. 2022.

[17] Farah Zhafira, D., & Rahayudi, B. "Analisis Sentimen Kebijakan Kampus Merdeka Menggunakan *Naive Bayes* dan Pembobotan *TF-IDF* Berdasarkan Komentar Pada *Youtube"*. Jurnal Sistem Informasi, Teknologi Informasi, Dan Edukasi Sistem Informasi (JUST-SI), 2(1), 55–63. 2021.