

OPTIMASI KINERJA METODE KLASIFIKASI DENGAN RANDOM SEARCH PADA ANALISIS SENTIMEN APLIKASI KORLANTAS POLRI

Aldino Marsel Pratama, Nuri Cahyono

Informatika, Universitas AMIKOM Yogyakarta

Jl. Ring Road Utara, Condong Catur, Sleman, Yogyakarta, Indonesia

aldinomarsel@students.amikom.ac.id

ABSTRAK

Kemajuan teknologi digital telah mendorong pengembangan aplikasi layanan publik, seperti Aplikasi Digital Korlantas Polri yang mempermudah proses perpanjangan SIM secara *online*. Namun, ulasan pengguna di *Google Play Store* menunjukkan berbagai masalah teknis, seperti kendala *login* dan kegagalan input data, yang memerlukan analisis mendalam terhadap persepsi pengguna. Penelitian ini bertujuan untuk menganalisis sentimen ulasan pengguna menggunakan metode klasifikasi *Support Vector Machine* (SVM) dan *K-Nearest Neighbors* (KNN) yang dioptimasi dengan *Random Search* serta ekstraksi fitur TF-IDF. Data penelitian terdiri dari 6.000 ulasan yang dikumpulkan pada periode 30 September 2024 hingga 20 November 2024. Data tersebut melalui tahapan *preprocessing* seperti *case folding*, *tokenisasi*, *stemming*, dan penyeimbangan data menggunakan SMOTE. Hasil penelitian menunjukkan bahwa optimasi *Random Search* berhasil meningkatkan kinerja model secara signifikan. Model KNN, yang dioptimasi dengan parameter *weights*: 'distance', *n_neighbors*: 5, dan *metric*: 'manhattan', mencapai akurasi 99,61%, dengan *precision*, *recall*, dan *F1-score* masing-masing mencapai 1.00 untuk semua kelas, kecuali untuk kategori negatif yang sedikit lebih rendah dengan *precision* 0.99. Sementara itu, SVM dengan parameter optimal (C: 100, *kernel*: 'linear', *gamma*: 0.01) menunjukkan akurasi 98%, dengan *precision* 0.97, *recall* 0.96, dan *F1-score* 0.96. Hasil penelitian ini diharapkan dapat memberikan wawasan kepada tim pengembang Digital Korlantas Polri untuk meningkatkan kualitas layanan berdasarkan analisis sentimen pengguna, serta menjadi referensi untuk penerapan metode optimasi lainnya seperti *Grid Search* atau penggunaan teknik ekstraksi fitur yang lebih canggih seperti *FastText* atau BERT.

Kata kunci : Analisis Sentimen, *K-Nearest Neighbors*, Korlantas, *Random Search*, *Support Vector Machine*.

1. PENDAHULUAN

Dalam era digital, perkembangan teknologi telah mengubah cara masyarakat mengakses layanan publik. Salah satu inovasi yang hadir adalah aplikasi Digital Korlantas Polri, yang memungkinkan perpanjangan Surat Izin Mengemudi (SIM) secara *online*. Aplikasi ini telah diunduh lebih dari satu juta kali dan menerima ribuan ulasan di *Google Play Store* [1], [2].

Pada situs resminya, aplikasi ini menjanjikan berbagai kemudahan, seperti pendaftaran dan ujian teori SIM secara *online*, perpanjangan SIM tanpa antrai, serta penggunaan *Digital ID* dan *Biometric Authentication* untuk meningkatkan keamanan. Selain itu, layanan ini diklaim dapat mempercepat proses administrasi, memudahkan pembayaran, dan menghilangkan kebutuhan untuk datang ke lokasi karena dokumen akan dikirim langsung ke rumah pengguna [3].

Namun, ulasan pengguna di *Google Play Store* menunjukkan pengalaman yang bertolak belakang dengan klaim resmi aplikasi. Banyak pengguna mengeluhkan berbagai kendala teknis, seperti kesulitan *login*, kegagalan *input data*, dan masalah performa *server*, yang justru memperumit proses administrasi dibandingkan mempermudahnya. Dengan adanya permasalahan tersebut, analisis sentimen diperlukan untuk mengolah ulasan pengguna guna memahami persepsi mereka terhadap aplikasi ini. Analisis ini bertujuan untuk mengidentifikasi dan mengelompokkan sentimen dalam ulasan apakah

bersifat positif, negatif, atau netral serta menggali wawasan dari data dalam jumlah besar [4].

Temuan ini juga menekankan pentingnya optimasi model dan pemahaman mendalam terhadap data dalam analisis sentimen.

Salah satu cara untuk meningkatkan performa metode klasifikasi adalah dengan mengoptimalkan *hyperparameter* menggunakan *Random Search*. *Random Search* terbukti lebih efisien dibandingkan *Grid Search* karena mampu mengalokasikan pencarian secara acak dalam ruang parameter, sehingga tidak terfokus pada eksplorasi dimensi parameter yang kurang relevan [5].

Dalam analisis sentimen, *Random Search* diterapkan untuk memilih kombinasi parameter yang optimal, memungkinkan model klasifikasi mencapai performa yang lebih baik secara efektif dan efisien dibandingkan metode optimasi konvensional [5].

Algoritma seperti *K-Nearest Neighbors* (KNN) dan *Support Vector Machine* (SVM) telah terbukti efektif dalam klasifikasi sentiment [1], [2], [6].

Dalam penelitian ini, *Random Search* digunakan untuk mengoptimalkan performa kedua algoritma tersebut dalam analisis sentimen ulasan aplikasi Digital Korlantas Polri. Selain itu, penelitian ini juga mengevaluasi perbandingan kinerja model sebelum dan sesudah optimasi guna mengukur efektivitas peningkatan yang dicapai. Hasil yang diperoleh diharapkan dapat meningkatkan akurasi analisis sentimen otomatis serta mendukung pemanfaatan

ulasan pengguna aplikasi layanan publik secara lebih optimal.

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Penelitian Kurnia Putri dkk. menganalisis sentimen terkait pengaruh brand ambassador terhadap preferensi pembelian konsumen menggunakan KNN dan SVM. Studi ini membantu perusahaan merancang strategi pemasaran yang lebih efektif [7].

Selanjutnya penelitian, Muhamad Fajri dan Aji Primajaya membandingkan *Grid Search* dan *Random Search* untuk optimasi SVM dalam klasifikasi. Hasilnya, *Random Search* 10x lebih cepat, sementara akurasi keduanya hampir sama (84%). Penggunaan memori tidak menunjukkan perbedaan signifikan, menjadikan *Random Search* lebih efisien untuk pencarian parameter [8].

Penelitian ini menunjukkan bahwa teknik optimasi *Random Search* berperan penting dalam meningkatkan performa klasifikasi sentimen. Dalam penelitian ini, algoritma SVM dan KNN dioptimasi menggunakan *Random Search* untuk menemukan kombinasi parameter terbaik. Pendekatan ini diharapkan dapat meningkatkan akurasi model dalam menganalisis sentimen ulasan aplikasi Digital Korlantas Polri.

2.2. Digital Korlantas Polri

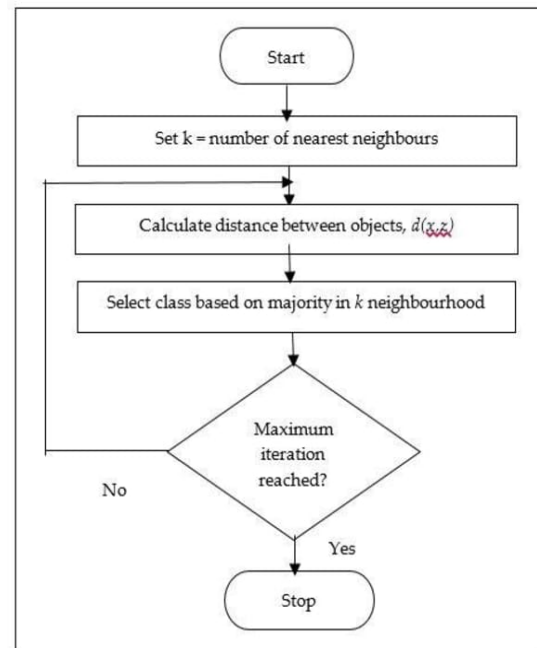
Aplikasi Digital Korlantas Polri merupakan platform yang dirancang untuk memudahkan masyarakat dalam memperbarui Surat Izin Mengemudi (SIM) secara *online*, diluncurkan oleh Korps Lalu Lintas POLRI dengan tujuan untuk meningkatkan aksesibilitas dan efisiensi layanan kepolisian di bidang lalu lintas. Aplikasi ini juga dilengkapi layanan SINGAL pembayaran pajak kendaraan bermotor (PKB) [1], [2].

2.3. Analisis Sentimen

Analisis sentimen adalah teknik NLP untuk mengidentifikasi sentimen dalam teks (positif, negatif, atau netral). Metode yang digunakan meliputi *machine learning*, *lexicon-based*, dan *hybrid*, dengan analisis pada level dokumen, kalimat, atau aspek. Teknik ini banyak diterapkan dalam *Customer Experience Management*, analisis media sosial, dan politik [9].

2.4. K-Nearest Neighbors (KNN)

K-Nearest Neighbors (KNN) adalah algoritma *machine learning* untuk klasifikasi berdasarkan kedekatan dengan data pelatihan. Algoritma ini sederhana namun rentan terhadap *noise*. Pemilihan nilai k yang optimal berpengaruh pada akurasi, di mana k lebih tinggi mengurangi *noise* tetapi dapat mengaburkan batas kategori. Optimasi parameter diperlukan untuk menentukan k terbaik [10].



Gambar 1. Contoh Flowchart KNN

Flowchart di atas menjelaskan proses klasifikasi KNN, dimulai dengan menentukan jumlah tetangga terdekat (k), menghitung jarak data uji ke data latih, lalu menetapkan kelas berdasarkan mayoritas k tetangga. Proses ini berulang hingga batas maksimum iterasi tercapai, kemudian hasil klasifikasi ditetapkan.

2.5. Support Vector Machine (SVM)

SVM adalah algoritma *supervised learning* yang mengklasifikasikan data dengan *hyperplane* sebagai pemisah dua kelas. Margin adalah jarak antara *hyperplane* dan titik terdekat (*support vector*). Tujuan utama SVM adalah menemukan *hyperplane* optimal dengan margin terlebar [11].

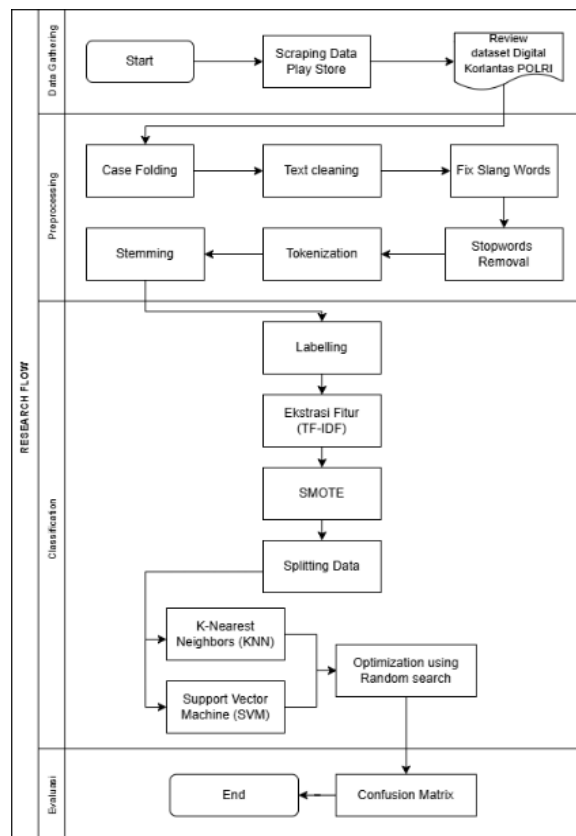
2.6. Random Search

Random Search adalah metode optimasi yang mengambil sampel parameter secara acak dan mengombinasikannya, fokus pada eksplorasi parameter yang paling berpengaruh. Tujuannya adalah menemukan kombinasi optimal tanpa pencarian menyeluruh di seluruh ruang parameter [8].

3. METODE PENELITIAN

3.1. Data Gathering

Dalam penelitian ini, data dikumpulkan menggunakan teknik *scraping* dengan memanfaatkan *library python*, yaitu *Google-Play-Scraper*. Proses pengumpulan data dilakukan melalui platform *Google Colab* menggunakan bahasa pemrograman *Python*. Pada tahap ini, data mentah yang dikumpulkan berupa ulasan pengguna yang diambil dari aplikasi Digital Korlantas POLRI di *Google Play Store*.



Gambar 2. Alur penelitian

3.2. Preprocessing

3.2.1. Case Folding

Pada tahap ini, semua huruf dalam data ulasan akan diubah menjadi huruf kecil (*lowercase*), yaitu dari 'A-Z' menjadi 'a-z', sehingga semua teks memiliki format penulisan yang seragam [14].

3.2.2. Text Cleaning

Langkah ini bertujuan untuk membersihkan data ulasan dari tautan, URL, simbol, dan karakter khusus yang tidak dibutuhkan, seperti tanda baca, emotikon, serta karakter khusus lainnya [15].

3.2.3. Slang Words Normalization

Tahap ini mengonversi kata *slang* atau singkatan ke dalam bentuk baku bahasa Indonesia untuk mempermudah analisis sentimen. Proses ini menggunakan kamus *slang* yang memetakan setiap kata *slang* ke bentuk standarnya [16].

3.2.4. Stopwords Removal

Tahap ini bertujuan untuk menghilangkan kata-kata umum yang sering muncul tetapi tidak memiliki makna signifikan dan tidak berguna dalam proses analisis sentimen, seperti: *adalah, aku, dan, atau, dia,* dan sebagainya [17].

3.2.5. Tokenizing

Langkah ini bertujuan untuk memisahkan ulasan menjadi kata atau frasa, sehingga memudahkan proses pengolahan data pada tahap berikutnya [18].

3.2.6. Stemming

Pada tahap ini, langkah yang dilakukan adalah menghapus imbuhan yang ada pada kata sehingga kata tersebut kembali ke bentuk dasarnya [19].

3.3. Labelling

Labelling mengategorikan ulasan berdasarkan rating pengguna di *Google Play Store* menjadi positif (4-5), netral (3), dan negatif (1-2). Label ini digunakan sebagai variabel target (*y*) dalam analisis sentimen, sementara ulasan yang telah diproses menjadi fitur (*X*) [20].

3.4. TF-IDF

Ekstraksi fitur mengubah teks menjadi numerik agar dapat diproses oleh *machine learning*. Metode TF-IDF menghitung frekuensi kata dalam dokumen serta keunikannya dalam kumpulan dokumen, menyoroti kata penting. Hasilnya digunakan sebagai fitur input (*X*) dalam pelatihan model [12].

3.5. SMOTE

Pada tahap ini, SMOTE diterapkan untuk menyeimbangkan jumlah sampel di kelas minoritas dengan cara menambahkan sampel hingga setara dengan kelas mayoritas. Melalui proses *oversampling* ini, data kelas minoritas diperbanyak, sehingga distribusi dataset menjadi lebih seimbang antara kelas minoritas dan kelas mayoritas [13].

3.6. Split Data

Pada tahap ini, dataset akan dibagi menjadi dua bagian, data latih (*training*) dan data uji (*testing*) dengan rasio pembagian umum seperti 80% untuk *training* dan 20% untuk *testing*. Data latih digunakan untuk melatih model yang akan dikembangkan, sedangkan data uji berfungsi untuk menguji kinerja model yang telah dibuat dalam proses klasifikasi.

3.7. Classification and Optimization

Data ulasan *Google Play Store* aplikasi Digital Korlantas POLRI yang telah melalui *preprocessing*, *labeling*, ekstraksi fitur, *split data*, dan SMOTE akan diklasifikasikan menggunakan *K-Nearest Neighbors* (KNN) dan *Support Vector Machine* (SVM). KNN menentukan label data uji berdasarkan mayoritas label dari tetangga terdekat, sementara SVM mencari *hyperplane* optimal untuk memisahkan kelas dengan margin maksimum. Model dilatih menggunakan data latih dan diuji dengan data uji untuk mengevaluasi performanya.

Agar kinerja model lebih optimal, *Random Search* digunakan untuk memilih kombinasi parameter terbaik secara acak dalam ruang pencarian. Evaluasi dilakukan berdasarkan akurasi, memilih kombinasi parameter yang menghasilkan performa terbaik.

3.8. Evaluation

Setelah klasifikasi dan optimasi, model KNN dan SVM dievaluasi menggunakan *confusion matrix*.

Evaluasi ini mengukur akurasi, presisi, recall, dan F1-score untuk menilai performa sebelum dan sesudah optimasi dengan *Random Search*.

Berikut persamaan rumus untuk setiap metrik evaluasi.

$$Accuracy = \frac{TP_A + TP_B + TP_C}{N} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1-Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (4)$$

Confusion matrix digunakan untuk mengevaluasi kesalahan klasifikasi dan mengidentifikasi kelemahan model dalam membedakan setiap kelas sentimen.

4. HASIL DAN PEMBAHASAN

4.1. Data Gathering

Penelitian ini mengumpulkan 6.000 ulasan aplikasi Digital Korlantas POLRI dari *Google Play Store* menggunakan *Google Colaboratory* dan *Google-Play-Scraper* pada 30 September 2024, mencakup periode 30 September – 20 November 2024. Data yang dikumpulkan mencakup isi ulasan (*content*) dan rating pengguna (*score*) untuk keperluan klasifikasi sentimen.

Tabel 1. Data Ulasan

	<i>content</i>	<i>Score</i>
0	😞	1
1	Sangat Baik pelayanan onlinenya	5
2	Lama proses perpanjangan SIM online 1 minggu lebih belum dikirim... Lebih cepat datang langsung offline dri pada online	1
3	Good	5
...	...	...
5999	Sekitar 2 minggu saya sudah peroleh SIM	5

Tabel ini menunjukkan beberapa contoh ulasan yang telah diklasifikasikan berdasarkan *score*, di mana nilai 1-2 menunjukkan sentimen negatif, 3 netral, dan 4-5 positif.

4.2. Preprocessing

4.2.1. Case Folding

Tabel 2 menunjukkan hasil *case folding*, di mana kolom *content* berisi teks asli, sementara kolom *clean_content* menampilkan teks yang telah dikonversi ke huruf kecil tanpa huruf kapital.

Tabel 2. *Case Folding*

	<i>content</i>	<i>Score</i>	<i>clean_content</i>
0	😞	1	😞
1	Sangat Baik pelayanan onlinenya	5	sangat baik pelayanan onlinenya
2	Lama proses perpanjangan SIM online 1 minggu lebih belum dikirim... Lebih cepat datang langsung offline dri pada online	1	lama proses perpanjangan sim online 1 minggu lebih belum di kirim... lebih cepat datang langsung offline dri pada online
3	Good	5	good
...	...	...	...
5999	Sekitar 2 minggu saya sudah peroleh SIM	5	sekitar 2 minggu saya sudah peroleh sim

4.2.2. Text Cleaning

Tabel 3. *Text Cleaning*

	<i>content</i>	<i>Score</i>	<i>clean_content</i>
0	😞	1	
1	sangat baik pelayanan onlinenya	5	sangat baik pelayanan onlinenya
2	lama proses perpanjangan sim online 1 minggu lebih belum di kirim... lebih cepat datang langsung offline dri pada online	1	lama proses perpanjangan sim online minggu lebih belum di kirim lebih cepat datang langsung offline dri pada online
3	good	5	good
...	...	...	...
5999	sekitar 2 minggu saya sudah peroleh sim	5	sekitar minggu saya sudah peroleh sim

Tabel 3 menampilkan ulasan sebelum dan sesudah *text cleaning*. Kolom *content* berisi teks asli, sementara *clean_content* menunjukkan ulasan yang telah dibersihkan dari tanda baca.

4.2.3. Slang Words Normalization

Tabel 4. *Slang Words Normalization*

	<i>content</i>	<i>Score</i>	<i>clean_content</i>
0		1	
1	sangat baik pelayanan onlinenya	5	sangat baik pelayanan onlinenya
2	lama proses perpanjangan sim online 1 minggu lebih belum di kirim... lebih cepat datang langsung offline dri pada online	1	lama proses perpanjangan sim online minggu lebih belum di kirim lebih cepat datang langsung offline dari pada online
3	good	5	bagus
...	...	...	...
5999	sekitar minggu saya sudah peroleh sim	5	sekitar minggu saya sudah peroleh sim

Tabel 4 menunjukkan teks sebelum dan sesudah *Slang Word Normalization*. Kolom *content* berisi teks asli, sedangkan *clean_content* menampilkan hasil normalisasi dengan kata-kata yang telah diubah ke bentuk baku.

4.2.4. Stopwords Removal

Tabel 5. *Stopwords Removal*

	<i>content</i>	<i>Score</i>	<i>clean_content</i>
0		1	
1	sangat baik pelayanan onlinenya	5	pelayanan onlinenya
2	lama proses perpanjang sim online minggu lebih belum di kirim lebih cepat datang langsung offline dari pada online	1	proses perpanjang sim online minggu kirim cepat langsung offline online
3	bagus	5	bagus
...	...	...	...
5999	sekitar minggu saya sudah peroleh sim	5	sekitar minggu peroleh sim

Tabel 5 menampilkan ulasan sebelum dan sesudah *Stopwords Removal*. Kolom *clean_content* menunjukkan hasil setelah kata imbuhan dan kata hubung dihapus dari *content*.

4.2.5. Tokenizing

Tabel 6. *Tokenizing*

	<i>content</i>	<i>Score</i>	<i>clean_content</i>
0		1	[]
1	pelayanan onlinenya	5	[pelayanan, onlinenya]
2	proses perpanjang sim online minggu kirim cepat langsung offline online	1	[proses, perpanjang, sim, online, minggu, kirim, cepat, langsung, offline, online]
3	bagus	5	[bagus]
...	...	...	...
5999	sekitar minggu peroleh sim	5	[minggu, peroleh, sim]

Tabel 6 menunjukkan teks sebelum dan sesudah *Tokenizing*. Kolom *clean_content* berisi teks yang telah dipisahkan menjadi satuan kata.

4.2.6. Stemming

Tabel 7. *Stemming*

	<i>content</i>	<i>Score</i>	<i>clean_content</i>
0	[]	1	[]
1	[pelayanan, onlinenya]	5	[layan, onlinenya]
2	[proses, perpanjang, sim, online, minggu, kirim, cepat, langsung, offline, online]	1	[proses, panjang, sim, online, minggu, kirim, cepat, langsung, offline, online]
3	[bagus]	5	[bagus]
...	...	...	...
5999	[minggu, peroleh, sim]	5	[minggu, oleh, sim]

Tabel 7 menampilkan teks sebelum dan sesudah *Stemming* menggunakan *Sastrawi*. Kolom *clean_content* berisi teks yang telah dikonversi ke bentuk kata dasar.

4.3. Labelling

Dalam penelitian ini, labelling dilakukan menggunakan fitur score dari ulasan *Google Play Store*. Tabel 8 menampilkan hasil labelling, di mana kolom *clean_content* berisi teks yang telah melalui *preprocessing*, sementara kolom *sentiment* menunjukkan kategori positif, negatif, atau netral. Dataset terdiri dari 2.263 ulasan positif, 3.423 negatif, dan 314 netral.

Tabel 8. *Labelling*

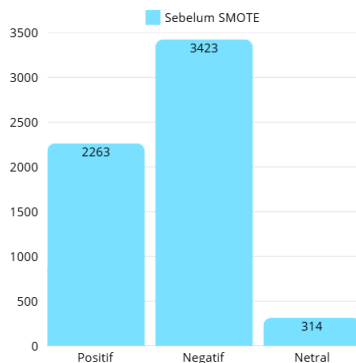
	<i>clean_content</i>	<i>score</i>	<i>sentiment</i>
0	[]	1	negatif
1	[layan, onlinenya]	5	positif
2	[proses, panjang, sim, online, minggu, kirim, cepat, langsung, offline, online]	1	negatif
3	[bagus]	5	positif
...	...	...	...
5999	[minggu, oleh, sim]	5	positif

4.4. TF-IDF

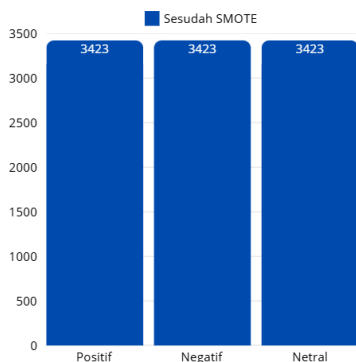
Setelah dilakukan *preprocessing* dan *labelling*, teks akan melalui tahap ekstraksi fitur menggunakan metode TF-IDF (*Term Frequency-Inverse Document Frequency*). TF-IDF menghasilkan representasi numerik untuk setiap kata berdasarkan frekuensi kemunculannya dalam dokumen dan dalam keseluruhan dataset. Proses ini bertujuan untuk mengubah teks menjadi format numerik yang dapat digunakan untuk analisis dan pemodelan lebih lanjut.

4.5. SMOTE

Karena data ulasan yang dikumpulkan tidak seimbang, sebagaimana terlihat pada Gambar 1 dari total 6.000 data terdapat 2.263 ulasan positif, 3.423 ulasan negatif, dan 314 ulasan netral. Oleh karena itu, perlu dilakukan *oversampling*. Dalam penelitian ini, peneliti menggunakan metode *Synthetic Minority Oversampling Technique* (SMOTE).



Gambar 3. Distribusi dataset sebelum SMOTE



Gambar 4. Distribusi dataset setelah SMOTE

Pada Gambar 2, persebaran data setelah proses SMOTE sudah seimbang, di mana kelas positif, negatif, dan netral masing-masing memiliki 3.423 data. Dengan demikian, total keseluruhan dataset setelah proses SMOTE berjumlah 10.269 data.

4.6. Split Data

Hasil dataset setelah dilakukan SMOTE selanjutnya akan dibagi menjadi data *train* dan data *test*, dengan perbandingan 80:20. Dari total dataset yang berjumlah 10.269, dibagi menjadi 8215 data training, dan 2054 data testing.

4.7. Classification and Optimization

Pada tahap ini, dilakukan penerapan dan optimasi dua algoritma klasifikasi, yaitu *K-Nearest Neighbors* (KNN) dan *Support Vector Machine* (SVM). Setiap algoritma diuji terlebih dahulu dengan parameter default, kemudian dioptimasi menggunakan *Random Search* untuk meningkatkan performa model.

4.7.1. K-Nearest Neighbors (KNN)

Model KNN awalnya diuji dengan konfigurasi default dan mencapai akurasi 70,69%. *Precision* pada kelas negatif cukup tinggi (98%), tetapi *recall* rendah (21%). Sebaliknya, kelas netral memiliki *recall* tinggi (98%) namun *precision* rendah (58%). Kelas positif menunjukkan keseimbangan yang lebih baik dengan *precision* 84% dan *recall* 93%. Untuk meningkatkan performa, dilakukan optimasi *hyperparameter* menggunakan *Random Search* dengan parameter antara lain *n_neighbors*: [3, 5, 7], *weights*: ['uniform', 'distance'], dan *metric*: ['euclidean', 'manhattan']. Hasil optimasi menunjukkan bahwa kombinasi parameter terbaik adalah *weights*: 'distance', *n_neighbors*: 5, dan *metric*: 'manhattan', yang meningkatkan akurasi menjadi 99,61%.

4.7.2. Support Vector Machine (SVM)

Model SVM awalnya diuji dengan konfigurasi default dan mencapai akurasi 97,6%, dengan *precision*, *recall*, dan *F1-score* rata-rata 0.98. Kelas negatif memiliki *precision* dan *recall* 0.99, kelas netral *precision* 1.00 tetapi *recall* lebih rendah (0.95), sementara kelas positif memperoleh *precision* 0.94 dan *recall* 0.99. Untuk meningkatkan performa, dilakukan optimasi *hyperparameter* menggunakan *Random Search* dengan parameter antara lain *C*: [1, 10, 100], *kernel*: ['linear'], dan *gamma*: [0.001, 0.01]. Hasil optimasi menunjukkan bahwa kombinasi parameter terbaik adalah *C*: 100, *kernel*: 'linear', *gamma*: 0.01, yang meningkatkan akurasi menjadi 98,7%.

4.8. Evaluation

Model dievaluasi menggunakan metrik akurasi, *precision*, *recall*, dan *F1-score*, serta *confusion matrix* untuk menganalisis pola kesalahan klasifikasi pada setiap model.

Tabel 9. Hasil evaluasi metrik klasifikasi dan *confusion matrix* (sebelum dan setelah optimasi *Random Search*)

Algoritma		Akurasi	Kelas	Precision	Recall	F1-Score	True Label			Predicted Label			
									Neg	Neu	Pos		
Metrik Evaluasi Sebelum Optimasi	KNN	70,69%	Neg	98%	21%	34%	Confusion Matrix Sebelum Optimasi	KNN	Neg	143	433	108	
			Neu	58%	98%	73%			Neu	0	669	16	
			Pos	84%	93%	88%			Pos	3	42	640	
	SVM	97,61%	Neg	92%	79%	85%		SVM	Neg	676	2	6	
			Neu	81%	92%	86%			Neu	0	651	34	
			Pos	89%	89%	89%			Pos	6	1	678	

Algoritma		Akurasi	Kelas	Precision	Recall	F1-Score	True Label			Predicted Label			
									Neg	Neu	Pos		
Metrik Evaluasi Setelah Optimasi	KNN	99,61%	Neg	100%	99%	99%	Confusion Matrix Setelah Optimasi	KNN	Neg	678	0	6	
			Neu	100%	100%	100%			Neu	1	683	1	
			Pos	99%	100%	99%			Pos	0	0	685	
	SVM	98,73%	Neg	95%	86%	90%		SVM	Neg	676	2	6	
			Neu	88%	97%	92%			Neu	0	668	17	
			Pos	95%	94%	94%			Pos	0	0	685	

Tabel 9 ,hasil evaluasi menunjukkan bahwa optimasi menggunakan *Random Search* meningkatkan akurasi, *precision*, *recall*, dan *F1-score* pada model KNN dan SVM. Sebelum optimasi, KNN memiliki akurasi 70,69%, dengan *precision* tinggi pada kelas negatif (98%) tetapi *recall* rendah (21%), sementara kelas netral memiliki *recall* tinggi (98%) namun *precision* rendah (58%). SVM mencatat akurasi 97,61%, dengan *precision* dan *recall* yang lebih seimbang dibandingkan KNN.

Setelah optimasi, performa kedua model meningkat signifikan. KNN mencapai akurasi 99,61%, dengan *precision* dan *recall* hampir sempurna di semua kelas (*precision* 99-100%, *recall* 99-100%), sementara SVM meningkat menjadi 98,73%, dengan *precision* dan *recall* yang lebih stabil dibandingkan sebelum optimasi. Analisis *confusion matrix* menunjukkan bahwa setelah optimasi, tingkat kesalahan klasifikasi menurun drastis, terutama pada kelas netral yang sebelumnya memiliki tingkat kesalahan tinggi.

Secara keseluruhan, optimasi dengan *Random Search* terbukti efektif dalam meningkatkan performa model klasifikasi. KNN mengalami peningkatan akurasi terbesar, sedangkan SVM tetap unggul dalam kestabilan prediksi di semua kelas.

5. KESIMPULAN DAN SARAN

Penelitian ini menyimpulkan bahwa metode KNN dan SVM dengan optimasi *Random Search* menunjukkan performa yang sangat baik dalam analisis sentimen ulasan aplikasi Digital Korlantas Polri. Dari 6.000 ulasan yang dikumpulkan dan diseimbangkan dengan SMOTE, KNN awalnya mencapai akurasi 70%, sedangkan SVM 97%. Setelah optimasi, akurasi KNN meningkat menjadi 99,61% dan SVM menjadi 98%. Untuk penelitian selanjutnya disarankan untuk mengeksplorasi metode lain seperti *Random Forest* atau *Naïve Bayes*, menggunakan model pralatih seperti BERT atau *FastText*, serta menerapkan optimasi *Grid Search* untuk meningkatkan efisiensi dan akurasi klasifikasi.

DAFTAR PUSTAKA

- [1] E. P. Sutrisno and S. Amini, "3 rd Seminar Nasional Mahasiswa Fakultas Teknologi Informasi (SENAFTI) 30 Agustus 2023-Jakarta," 2023.
- [2] E. R. Kaburuan and N. R. Setiawan, "Sentimen Analisis Review Aplikasi Digital Korlantas Pada Google Play Store Menggunakan Metode SVM," *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, vol. 12, no. 1, pp. 105–116, Mar. 2023, doi: 10.32736/sisfokom.v12i1.1614.
- [3] Kepolisian Republik Indonesia, "Digital Korlantas Polri," <https://www.digitalkorlantas.id/>. Accessed: Feb. 10, 2025. [Online]. Available: <https://www.digitalkorlantas.id/>
- [4] K. Cindy Astuti, A. Firmansyah, A. Riyadi, and U. Pelita Bangsa Bekasi, "Implementasi Text Mining untuk Analisis Sentimen Masyarakat terhadap Ulasan Aplikasi Digital Korlantas Polri pada Google Play Store," *Remik: Riset dan E-Jurnal Manajemen Informatika Komputer*, vol. 8, no. 1, 2024, doi: 10.33395/remik.v8i1.13421.
- [5] Fatimah Rahmadayana and Yuliant Sibaroni, "Sentiment Analysis of Work from Home Activity using SVM with Randomized Search Optimization," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 5, pp. 936–942, Oct. 2021, doi: 10.29207/resti.v5i5.3457.
- [6] A. S. Pamungkas and N. Cahyono, "Analisis Sentimen Review ChatGPT di Play Store menggunakan Support Vector Machine dan K-Nearest Neighbor," *Edumatic: Jurnal Pendidikan Informatika*, vol. 8, no. 1, pp. 1–10, Jun. 2024, doi: 10.29408/edumatic.v8i1.24114.
- [7] N. K. Putri, A. V. Vitianingsih, S. Kacung, A. L. Maukar, and V. Yasin, "Sentiment Analysis of Brand Ambassador Influence on Product Buyer Interest Using KNN and SVM," *Indonesian Journal of Artificial Intelligence and Data Mining*, vol. 7, no. 2, p. 327, Apr. 2024, doi: 10.24014/ijaidm.v7i2.29469.
- [8] M. Fajri and A. Primajaya, "Komparasi Teknik Hyperparameter Optimization pada SVM untuk Permasalahan Klasifikasi dengan Menggunakan Grid Search dan Random Search," 2023. [Online]. Available: <http://jurnal.polibatam.ac.id/index.php/JAIC>
- [9] Khadijah, U.N.; Cahyono, N.; Abstrak, I.A. Analisis Topic Modelling Pariwisata Yogyakarta Menggunakan Latent Dirichlet Allocation (LDA). *Indones. J. Comput. Sci.* 2024, 13, 6075–6086

- [10] H. Wisnu, M. Afif, and Y. Ruldevyani, "Sentiment analysis on customer satisfaction of digital payment in Indonesia: A comparative study using KNN and Naïve Bayes," in *Journal of Physics: Conference Series*, Institute of Physics Publishing, Feb. 2020. doi: 10.1088/1742-6596/1444/1/012034.
- [11] J. Ipmawati, S. Saifulloh, and K. Kusnawi, "Analisis Sentimen Tempat Wisata Berdasarkan Ulasan pada Google Maps Menggunakan Algoritma Support Vector Machine," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 1, pp. 247–256, Jan. 2024, doi: 10.57152/malcom.v4i1.1066.
- [12] A. Oktavia Praneswara and N. Cahyono, "Analisis Sentimen Ulasan Aplikasi TikTok Shop Seller Center di Google Playstore Menggunakan Algoritma Naive Bayes," *Indonesian Journal of Computer Science Attribution*, vol. 12, no. 6, p. 3925, 2023.
- [13] B. Wicaksono and N. Cahyono, "ANALISIS SENTIMEN KOMENTAR INSTAGRAM PADA PROGRAM KAMPUS MERDEKA DENGAN ALGORITMA NAIVE BAYES DAN DECISION TREE," 2024. [Online].
- [14] A. R. D. Astuti and N. Cahyono, "Analisis Topic Modelling Persepsi Pengguna Internet Menggunakan Metode Latent Dirichlet Allocation," *Indones. J. Comput. Sci.*, vol. 12, no. 1, pp. 326–334, 2023, doi: 10.33022/ijcs.v12i1.3155.
- [15] D. Setiyawati and N. Cahyono, "Analisa Sentimen Pengguna Sosial Media Twitter Terhadap Perokok di Indonesia," *Indonesian Journal of Computer Science*, vol. 12, no. 1, pp. 262–272, 2023. DOI: <https://doi.org/10.33022/ijcs.v12i1.3154>.
- [16] F. Baehaqi and N. Cahyono, "Analisis Sentimen Terhadap Cyberbullying Pada Komentar Di Instagram Menggunakan Algoritma Naive Bayes," *Indonesia. J. Comput. Sci.*, vol. 13, no. 1, pp. 1051–1063, 2024, doi: 10.33022/ijcs.v13i1.3301.
- [17] R. Rakarahayu Putri and N. Cahyono, "Analisis Sentimen Komentar Masyarakat Terhadap Pelayanan Publik Pemerintah Dki Jakarta Dengan Algoritma Super Vector Machine Dan Naive Bayes," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 2, pp. 2363–2371, 2024, doi: 10.36040/jati.v8i2.9472.
- [18] R. G. A. Pratama and N. Cahyono, "Comparison of Deep Learning Methods on Sentiment Analysis Using Word Embedding," *JITK*, vol. 10, no. 1, pp. 1–8, Jul. 2024. DOI: <https://doi.org/10.33480/jitk.v10i1.5280>
- [19] G. A. Joni, N. Cahyono, A. Baita, and N. Aini, "Rangka Esaf Honda Menggunakan Algoritma Naive Bayes Dan Super Vector Machine," vol. 8, no. 5, pp. 10879–10885, 2024. DOI: <https://doi.org/10.36040/jati.v8i5.11164>.
- [20] A. N. Safira, E. Pujastuti, H. Hanafi, B. Setiaji, D. Prabowo, and N. Cahyono, "Sentiment Analysis to Find Out Positive or Negative Opinions on Ride Hailing Application," in 2023 International Conference on Informatics, Multimedia, Cyber and Informations System (ICIMCIS), Jakarta Selatan, Indonesia, 2023, pp. 341–345, doi: 10.1109/ICIMCIS60089.2023.10349083.