

ANALISIS SENTIMEN ULASAN PENGGUNA APLIKASI MOBILE SP4N-LAPOR! DENGAN PENDEKATAN MACHINE LEARNING

Yeni Sulistiowati¹, Bagus Jati Santoso²

¹Manajemen Teknologi, Sekolah Interdisipliner Manajemen dan Teknologi, Institut Teknologi Sepuluh November

²Teknik Informatika, Fakultas Teknologi Elektro dan Informatika Cerdas, Institut Teknologi Sepuluh November

¹6032231247@student.its.ac.id, ²bagus@if.its.ac.id

Abstrak

Kemajuan teknologi informasi di Indonesia telah mendorong peningkatan akses internet, yang kini mencapai 79,5% dari total populasi. Kondisi ini menciptakan peluang besar bagi digitalisasi pelayanan publik, termasuk penerapan Sistem Pengelolaan Pengaduan Pelayanan Publik Nasional (SP4N-LAPOR!). Aplikasi ini memungkinkan masyarakat untuk menyampaikan pengaduan dengan cepat dan transparan melalui berbagai platform, termasuk aplikasi *mobile*. Semakin populernya perangkat *mobile* semakin memperkuat peran SP4N-LAPOR! dalam mendorong partisipasi masyarakat dalam pengawasan pelayanan publik. Namun, sebagian besar penelitian sebelumnya masih terbatas pada analisis data kualitatif dan kuantitatif deskriptif, sehingga pengalaman serta perspektif pengguna sebagai elemen utama dalam aplikasi ini belum tergambar secara komprehensif. Untuk mengatasi keterbatasan tersebut, penelitian ini mengusulkan penggunaan analisis sentimen terhadap ulasan masyarakat yang diberikan melalui aplikasi *mobile* berbasis Android. Pendekatan ini bertujuan untuk mengidentifikasi permasalahan teknis serta memberikan gambaran mengenai kualitas pelayanan yang disediakan oleh penyelenggara. Analisis dilakukan menggunakan teknik *Machine Learning* dengan tiga model utama: *Naive Bayes* (NB), *Support Vector Machine* (SVM), dan kombinasi keduanya (NBSVM) dengan tambahan fitur ekspansi *N-gram*. Hasil pengujian menunjukkan bahwa model NBSVM memiliki performa terbaik dengan *G-Mean* sebesar 0.8451, *Sensitivity* sebesar 0.8227 dan *F2 Score* sebesar 0.8215, mengungguli model NB dan SVM. Secara praktis, hasil klasifikasi sentimen dapat dimanfaatkan sebagai sistem peringatan dini (*early warning system*) untuk memprioritaskan aduan dengan sentimen negatif, sekaligus menjadi alat evaluasi bagi penyelenggara dalam meningkatkan kualitas pelayanan. Dengan demikian, penelitian ini mendukung pengambilan keputusan berbasis data, serta memperkuat peran Ombudsman RI sebagai pengawas pelaksanaan SP4N-LAPOR!.

Kata kunci: Analisis Sentimen, *Naive Bayes*, NBSVM, *N-gram*, *Support Vector Machine*

1. Pendahuluan

Kemajuan teknologi digital telah mendorong pemerintah untuk mengembangkan kanal pengaduan berbasis *mobile* guna meningkatkan aksesibilitas dan efisiensi pelayanan publik. Aplikasi *mobile* SP4N-LAPOR! hadir sebagai bagian dari transformasi digital dalam pengelolaan pengaduan masyarakat yang pelaksanaannya diawasi oleh Ombudsman RI sebagaimana tertuang pada Permen PAN & RB No. 46 Tentang Road Map Sistem Pengelolaan Pengaduan Pelayanan Publik Nasional Tahun 2020-2024 (2020). Menurut penelitian Pusvita & Muttaqin (2023), meskipun *website* masih menjadi kanal terbanyak yang digunakan dalam pengaduan, aplikasi *mobile* menawarkan fleksibilitas, kemudahan akses, serta interaksi yang lebih cepat dan responsif, sehingga semakin relevan dalam mendorong partisipasi masyarakat dalam pengawasan pelayanan publik.

Namun, meskipun sistem ini telah dirancang sesuai regulasi, penelitian Premaiswari & Made (2024) menunjukkan bahwa implementasinya masih

menghadapi berbagai kendala, seperti kurangnya komitmen, koordinasi yang belum optimal, keterbatasan kapasitas teknis dan operasional, tantangan dalam aspek teknologi, serta minimnya keterlibatan masyarakat. Permasalahan ini menghambat efektivitas sistem dalam menangani keluhan publik secara transparan dan responsif.

Sebagian besar penelitian terdahulu menggunakan pendekatan deskriptif kualitatif, yang berfokus pada pemetaan kualitas layanan berdasarkan indikator tetap, seperti *E-GovQual* dalam penelitian Wildhani et al. (2023). Namun, pendekatan ini memiliki keterbatasan dalam menangkap opini pengguna secara *real-time* dan mengidentifikasi masalah teknis yang muncul secara dinamis. Oleh karena itu, analisis sentimen berbasis *machine learning* menjadi pendekatan yang lebih fleksibel dan otomatis dalam mengekstrak opini masyarakat dari ulasan pengguna. Teknik ini tidak hanya memungkinkan identifikasi pola sentimen publik terhadap pelayanan SP4N-LAPOR!, tetapi juga mendeteksi potensi kendala teknis yang mungkin

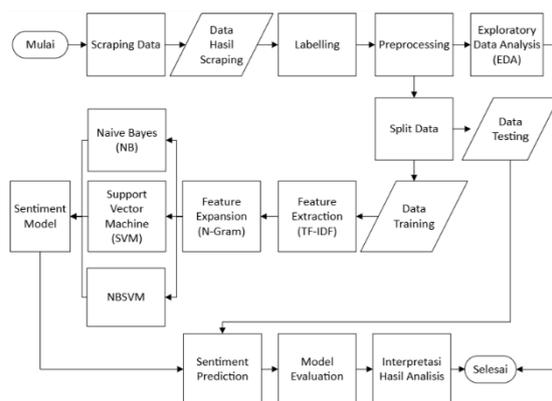
tidak terdeteksi melalui metode kualitatif konvensional.

Penelitian oleh A'la (2022) membahas tentang analisis sentimen ulasan berbahasa Indonesia, dan telah mengkaji metode *machine learning* seperti *Naive Bayes* (NB) dan *Support Vector Machine* (SVM) yang menunjukkan keefektifannya dalam menangani hal tersebut. Selain itu menurut penelitian oleh Herlinawati et al. (2020); Hermanto et al. (2020); Hakim et al. (2021); Mustakim & Priyanta (2022); Anreaja et al. (2022) mengatakan bahwa tingkat akurasi yang didapatkan algoritma *Support Vector Machine* (SVM) lebih unggul dibandingkan *Naive Bayes* (NB). Sementara itu, penelitian oleh Wang & Manning (2012); Muhammad (2019); Lestari & Setiawan (2022) mengatakan bahwa penggunaan kombinasi algoritma antara *Naive Bayes* (NB) dan *Support Vector Machine* (SVM) atau dikenal dengan NBSVM dan menghasilkan performa yang lebih baik.

Berdasarkan uraian tersebut, penelitian ini melakukan analisis sentimen menggunakan pendekatan *machine learning* untuk memahami konteks ulasan pada aplikasi mobile SP4N-LAPOR!. Algoritma yang digunakan meliputi *Naive Bayes* (NB), *Support Vector Machine* (SVM), serta kombinasi keduanya (NBSVM) dengan tambahan fitur ekspansi *N-Gram* kombinasi *unigram* dan *bigram*. Meskipun algoritma tersebut sudah banyak digunakan, namun dalam penelitian ini dirancang untuk menangkap konteks bahasa dan istilah teknis dalam data ulasan aplikasi pengelolaan pengaduan pelayanan publik yang mana masih jarang dikaji dalam literatur sebelumnya. Selain itu, adanya tambahan fitur kombinasi *unigram* dan *bigram*, memungkinkan model untuk menangkap konteks lokal seperti frasa-frasa khas pelayanan publik, seperti “tidak ditanggapi”, “proses lama”, dan “pelayanan cepat”. Strategi tersebut dilakukan dengan harapan dapat meningkatkan sensitivitas model dalam memahami makna sentimen yang tersirat maupun eksplisit dalam aduan masyarakat. Hasil penelitian ini diharapkan dapat menjadi alat evaluasi yang berguna bagi penyelenggara, baik dari sisi teknis penggunaan aplikasi maupun peningkatan kualitas pelayanan yang menjadi inti dari aspirasi dan pengaduan masyarakat.

2. Metode

Tahapan penelitian ini terdiri atas pengumpulan data, *labelling* data, *preprocessing* data, *Exploratory Data Analysis*, *split* data menjadi data *training* dan *testing*, *feature extraction*, *feature expansion*, klasifikasi dengan ketiga algoritma (NB, SVM, dan NBSVM), evaluasi model, serta interpretasi hasil analisis yang mencakup beberapa temuan yang akan disampaikan terkait gambaran atas konteks ulasan pengguna pada aplikasi *mobile* SP4N-LAPOR!. Untuk lebih jelasnya, alur penelitian dapat dilihat pada Gambar 1.



Gambar 1. Alur Penelitian

2.1 Pengumpulan Data

Data yang digunakan dalam penelitian ini yaitu data ulasan pengguna aplikasi *mobile* SP4N-LAPOR! dengan pengaturan bahasa dan negara Indonesia. Data diambil sejak dari awal aplikasi diluncurkan hingga tanggal 31 Desember 2024. Pengumpulan data dilakukan dengan cara *scraping* menggunakan *library google-play-scrapers*, *pandas*, dan *numpy*. Data yang berhasil dihimpun akan disimpan dalam bentuk file dengan format csv untuk diolah lebih lanjut.

2.2 Labelling Data

Dalam analisis sentimen, proses pelabelan data tidak selalu terbatas pada kategori positif, netral, dan negatif. Menurut penelitian Prabowo & Thelwall (2009), sentimen juga dapat diklasifikasikan dalam berbagai bentuk lain, seperti setuju atau tidak setuju, baik atau buruk, atau kategori lain yang sesuai dengan konteks. Tujuan utama dari pelabelan ini yaitu untuk mengelompokkan data berdasarkan sentimen tertentu sebelum memasuki tahap pelatihan, sehingga model dapat mengenali pola sentimen dengan lebih baik. Dalam penelitian ini, pelabelan dilakukan berdasarkan *score based (rating)* ulasan. Ulasan yang memiliki *rating* 1, 2, dan 3 dikategorikan sebagai negatif, sedangkan yang memiliki *rating* 4 dan 5 diberi label non-negatif.

2.3 Preprocessing Data

Pemrosesan teks merupakan tahap krusial dalam penambahan teks. Proses ini bertujuan untuk mengubah teks yang tidak terstruktur menjadi format yang lebih terstruktur, sehingga dapat dianalisis dengan lebih efektif.

Tahapan pertama yang dilakukan yaitu *case folding*. Menurut Saputro & Hermawan (2021) dalam penelitiannya, pada tahapan ini dilakukan perubahan pada huruf dalam teks menjadi huruf kecil agar penggunaan huruf menjadi standar. Selanjutnya yaitu tahapan *symbol removal*, yang merupakan proses menghapus tanda baca, karakter khusus maupun

angka yang tidak berarti. Tahap berikutnya melakukan *word normalization*, untuk mengubah bahasa tidak standar (*slang*) menjadi baku. Setelah itu dilakukan *tokenization*, yang menurut penelitian Palomino & Aider (2022) yaitu merupakan tahapan membagi sepotong teks menjadi bagian-bagian yang disebut token. Berikutnya yaitu dilakukan tahapan *stopwords*. Menurut Xu et al. (2024) pada tahapan ini dilakukan penghapusan kata-kata yang sering muncul namun tidak membantu dalam memahami teks. Contoh *stopwords* antara lain “yang”, “akan”, dll. Tahapan terakhir dari *preprocessing data* yaitu *stemming*. Menurut Saputro & Hermawan (2021), tahapan ini dilakukan untuk mengubah kata menjadi bentuk dasar.

2.4 Exploratory Data Analysis

Setelah tahapan *preprocessing data* selesai, maka selanjutnya dilakukan *Exploratory Data Analysis (EDA)* untuk memahami pola sentimen dalam data. Distribusi sentimen divisualisasikan untuk melihat proporsi kategori negatif dan non-negatif. *Word cloud* digunakan untuk mengidentifikasi kata-kata dominan dalam setiap sentimen. Menurut Nurandi (2015), semakin sering satu kata digunakan, maka semakin besar pula ukuran kata tersebut ditampilkan dalam *word cloud*. Selain itu, EDA juga mengungkap potensi ketidakseimbangan data, serta memastikan teks telah siap untuk pemodelan analisis sentimen. Hasil dari tahap ini memberikan wawasan mendalam yang mendukung pengembangan model yang lebih akurat dan optimal.

2.5 Split Data

Tahapan berikutnya yaitu melakukan *split data* (pembagian data). Data ulasan hasil dari *preprocessing* dibagi menjadi dua, yaitu dengan proporsi sebesar 80% data *training* (latih) dan 20% data *testing* (uji). *Proses split data* bertujuan untuk memastikan model dapat belajar pola dari data latih dan dievaluasi kinerjanya menggunakan data uji yang belum pernah dilihat sebelumnya.

Menurut Nawaz (2023), untuk menjaga keseimbangan distribusi sentimen dalam kedua subset, pembagian dilakukan secara stratifikasi berdasarkan kategori sentimen yang ada. Dengan pendekatan ini, model dapat diuji secara objektif dan menghasilkan klasifikasi sentimen yang lebih akurat serta memiliki kemampuan generalisasi yang lebih baik.

2.6 Feature Extraction

Dalam penelitian ini, *feature extraction* dilakukan menggunakan *Term Frequency-Inverse Document Frequency (TF-IDF)* untuk menentukan

seberapa penting suatu kata dalam mewakili sebuah kalimat. Menurut Nawaz (2023), TF-IDF memberikan bobot yang lebih tinggi pada kata-kata yang sering muncul dalam suatu dokumen tetapi jarang muncul di seluruh kumpulan data, sehingga membantu dalam mengidentifikasi kata-kata yang paling relevan untuk analisis sentimen.

2.7 Feature Expansion

Selain itu, untuk menangkap informasi terkait urutan kata dalam kalimat, dilakukan *feature expansion* menggunakan *N-Gram*. Menurut Das et al. (2020), *N-Gram* merepresentasikan sekumpulan 'N' kata yang berurutan. *Unigram* merepresentasikan setiap kata secara individual, sedangkan *bigram* mempertimbangkan kombinasi dua kata berurutan, sehingga dapat menangkap hubungan antar kata yang lebih baik. Dalam penelitian ini penggunaan *N-Gram* dibatasi pada kombinasi *unigram* dan *bigram*. Pembatasan ini sejalan dengan penelitian Adyatma et al. (2023) yang menyatakan bahwa penggunaan kombinasi *unigram* dan *bigram* dapat meningkatkan akurasi dari model klasifikasi.

2.8 Klasifikasi Sentimen

Pada penelitian ini terdapat beberapa skenario algoritma *machine learning* yang digunakan untuk klasifikasi. Diantaranya yaitu *Naive Bayes (NB)*, *Support Vector Machine (SVM)* dan kombinasi keduanya NBSVM. Skenario pertama yang digunakan yaitu *Multinomial Naive Bayes (MNB)*. Menurut Abbas et al. (2019) MNB merupakan salah satu jenis algoritma NB yang dirancang untuk klasifikasi. Skenario kedua dalam penelitian ini yaitu menggunakan *Support Vector Machine (SVM)*. Menurut Faisal (2019), SVM merupakan algoritma yang bekerja dengan mencari *hyperplane* optimal yang dapat memisahkan kelas-kelas dalam data dengan margin terbesar. Dalam penelitian ini model yang digunakan adalah SVM linier karena ukuran data yang tidak terlalu besar. Skenario berikutnya menggunakan NBSVM yang merupakan kombinasi antara NB dan SVM. Dalam penelitian Wang & Manning (2012); Muhammad (2019), NBSVM diklaim memiliki kinerja yang sangat baik untuk menangani cuplikan dokumen atau dokumen berdurasi penuh dan memiliki tingkat akurasi yang baik. Pendekatan ini memanfaatkan estimasi probabilitas dari *Naive Bayes* untuk mengubah representasi fitur sebelum diterapkan ke model SVM dengan kernel linier. Dalam NBSVM, rasio log-probabilitas kelas dihitung terlebih dahulu berdasarkan distribusi kata menggunakan NB, kemudian hasil tersebut digunakan sebagai input bagi SVM linier untuk menentukan *hyperline* yang optimal.

2.9 Evaluasi Model

Setelah dilakukan klasifikasi model sentimen, maka tahapan berikutnya yaitu mengevaluasi model. Evaluasi model ini dapat dilakukan dengan menggunakan *confusion matrix*. *Confusion matrix* menurut Alqaryouti et al. (2024) digunakan untuk mengukur dan menghitung *precision score*, *accuracy*, *recall*, dan *F1-measurement*. Namun menurut Bekkar et al. (2013); Wang et al. (2021) dalam kasus data *imbalanced*, pengukuran ketepatan klasifikasi yang digunakan adalah *G-mean*, *Specificity* dan nilai *Area Under Curve* (AUC) yang merupakan indikator performansi kurva *Receiver Operating Characteristic* (ROC) yang dapat meringkas kinerja *classifier* menjadi satu nilai. Sementara itu, pada penelitian Prasetiyo et al. (2021); Truong et al. (2024) menyebutkan bahwa performansi model dengan data *imbalanced* juga dapat menggunakan *F2 Score* jika berfokus untuk menghindari *false negative* karena *recall* menjadi lebih penting. Sebagai contoh pada penelitian Truong et al. (2024), *recall* menjadi lebih penting karena menyangkut prediksi kejadian stroke dan perdarahan pada pasien, yang apabila terjadi kegagalan dalam mengidentifikasi pasien (kesalahan klasifikasi) maka dapat berdampak fatal. Hal ini sejalan dengan yang dilakukan dalam penelitian yang dilakukan, bahwa *recall* menjadi lebih penting, karena jika terjadi kegagalan dalam identifikasi sentimen negatif, maka ulasan yang sebenarnya membutuhkan perhatian dan tindak lanjut dari penyelenggara pelayanan dapat terabaikan, sehingga berdampak pada kualitas layanan dan kepuasan pengguna pelayanan.

3. Hasil dan Pembahasan

Total data ulasan yang berhasil didapatkan pada penelitian ini sebanyak 1.687. Setelah dilakukan *labelling* data berdasarkan *rating* ulasan secara keseluruhan, didapatkan 984 ulasan atau 58% data memiliki sentimen negatif, sementara itu 703 ulasan atau 42% data memiliki sentimen non-negatif. Hal ini menunjukkan adanya kondisi *imbalanced* pada data karena ukuran kategori sentimen yang berbeda antara label negatif dan non-negatif. Sehingga dalam penelitian ini, pengukuran ketepatan klasifikasi dalam tahapan evaluasi model yang digunakan adalah *G-Mean*, *Specificity*, *AUC*, dan *F2 Score*.

Tahapan berikutnya yaitu *preprocessing* data dengan tujuan untuk mengubah data ulasan yang tidak terstruktur menjadi data yang lebih terstruktur agar dapat dilakukan analisis lebih lanjut. Contoh data hasil *preprocessing* data ulasan disajikan pada Tabel 1.

Tabel 1 Tabel Contoh Data Hasil *Preprocessing* Data

No	Ulasan	Hasil <i>Preprocessing</i>
1	aplikasi bosok lemot asu	"bosok", "anjing"

2	semangat dari pembuatan "semangat", aplikasinya bagus tapi klo gak "aplikasi", bisa login dan loading terus apa "masuk", bisa dibilang bagus?	"buat", "bagus", "loading", "bilang", "bagus"
---	--	--

Data hasil *preprocessing* kemudian dianalisis lebih lanjut menggunakan *word cloud*. Berdasarkan hasil *word cloud* dapat dilihat fitur apa saja yang paling sering muncul pada masing-masing kategori sentimen. *Word cloud* untuk sentimen non-negatif dapat dilihat pada Gambar 2, sedangkan untuk sentimen negatif dapat dilihat pada Gambar 3.



Gambar 2. *Word cloud* Sentimen Non-Negatif



Gambar 3. *Word cloud* Sentimen Negatif

Berdasarkan Gambar 2, terlihat bahwa kata-kata seperti "lapor", "masyarakat", dan "moga" sering muncul, yang menunjukkan bahwa aplikasi ini dimanfaatkan sesuai dengan fungsinya sebagai media untuk menyampaikan keluhan atau aspirasi publik. Sementara itu, kata-kata seperti "bantu", "mantap", "bagus", "cepat", "mudah", dan "manfaat" yang mendominasi dalam sentimen non-negatif menunjukkan bahwa para pengguna mengapresiasi keunggulan aplikasi ini, terutama dalam hal kecepatan respons terhadap laporan serta kemudahan penggunaannya.

Sementara itu pada Gambar 3, kata "lapor" juga sering muncul, mencerminkan adanya laporan yang masuk ke aplikasi serta keluhan dari pengguna yang memiliki sentimen negatif. Selain itu, kata-kata seperti "server ganggu", "ganggu", "masuk", "aplikasi", "daftar", "tanggap", "respon", "perintah", dan "verifikasi" muncul dengan frekuensi tinggi, mengindikasikan permasalahan utama yang dikeluhkan pengguna, seperti gangguan teknis pada server, kendala dalam proses pendaftaran atau

verifikasi, serta kegagalan dalam mengirim laporan. Kemunculan kata "adu", "tindak", dan "tindak lanjut" semakin menegaskan sentimen negatif seperti pengguna merasa bahwa aplikasi belum memberikan solusi yang memadai, atau respons yang cepat terhadap permasalahan yang mereka hadapi.

Pada tahapan selanjutnya, data ulasan dibagi menjadi dua bagian dengan proporsi 80% untuk data *training* (sejumlah 1.349 data) dan 20% untuk data *testing* (sejumlah 338 data) menggunakan metode *stratified split*, agar distribusi label sentimen negatif (58%) dan non-negatif (42%) tetap seimbang pada kedua subset. Pendekatan ini memastikan bahwa evaluasi model mencerminkan performa yang representatif terhadap keseluruhan data. Proses *feature extraction* dilakukan dengan metode TF-IDF untuk mengukur relevansi kata dalam kalimat, serta ditambahkan ekspansi fitur *N-Gram* terbatas pada kombinasi *unigram* dan *bigram*.

Berikutnya dilakukan klasifikasi sentimen pada data *training* dengan menggunakan ketiga algoritma yang diusulkan dalam penelitian ini. Hasil evaluasi atas ketiga algoritma klasifikasi dapat dilihat pada Tabel 2. Pada tabel tersebut terlihat bahwa meskipun model NB unggul dalam mengenali kategori sentimen negatif, kemampuannya dalam mengidentifikasi kategori sentimen non-negatif masih terbatas. Di sisi lain, model SVM memiliki performa yang lebih seimbang dibandingkan NB, tetapi tetap mengalami kendala dalam mendeteksi kategori sentimen non-negatif secara optimal.

Tabel 2 Tabel Perbandingan Hasil Evaluasi Model Naïve Bayes (NB), Support Vector Machine (SVM), dan NBSVM

Keterangan	Model		
	NB	SVM	NBSVM
<i>Specificity</i>	0.9746	0.9340	0.8680
<i>Sensitivity</i>	0.5745	0.6312	0.8227
<i>G-Mean</i>	0.7483	0.7678	0.8451
<i>AUC</i>	0.9100	0.8840	0.8884
<i>F2 Score</i>	0.6231	0.6682	0.8215
<i>Accuracy</i>	0.8100	0.8100	0.8500

Sementara itu, model gabungan NBSVM menunjukkan hasil yang lebih unggul dalam mengklasifikasikan kedua kategori sentimen. Model ini mencatat nilai *G-Mean* 0.8451, *Sensitivity* sebesar 0.8227 dan *F2 Score* yang cukup tinggi sebesar 0.8215. Selain itu, terlihat juga bahwa akurasi NBSVM juga lebih baik dibandingkan dengan NB dan SVM. Hal ini mengindikasikan kinerja keseluruhan model NBSVM lebih optimal.

Keunggulan model NBSVM tersebut dikarenakan menggabungkan kekuatan generalisasi probabilistik dari Naive Bayes dan kemampuan pemisahan kelas yang kuat dari SVM. Pendekatan NBSVM dalam penelitian ini efektif dalam mengatasi tantangan pada data teks yang cenderung tidak seimbang dan memiliki variasi ekspresi linguistik, seperti yang ditemukan dalam ulasan pengguna

SP4N-LAPOR!. Secara praktis, model ini mampu secara lebih akurat mengidentifikasi aduan dengan muatan negatif yang membutuhkan perhatian lebih lanjut, menjadikannya relevan untuk diintegrasikan dalam sistem deteksi dini (*early warning system*).

Sebagai pelengkap evaluasi kuantitatif, dilakukan analisis terhadap beberapa contoh ulasan yang berhasil dan tidak berhasil diklasifikasikan oleh model. Pada sejumlah aduan bernada negatif, seperti "jelek sekali kenapa aplikasi yang dibuat negara kok seperti main susah akses susah masuk susah kirim..." dan "tidak bisa digunakan selalu ada pemberitahuan internal server errors", seluruh model mampu mengenali ekspresi ketidakpuasan secara akurat. Kata-kata seperti "jelek", "susah akses", dan "server errors" menjadi indikator linguistik yang kuat terhadap sentimen negatif. Namun, terdapat pula kalimat dengan makna yang kurang eksplisit sehingga menimbulkan tantangan dalam klasifikasi, seperti "masih mencoba", yang oleh model NBSVM diinterpretasikan sebagai sentimen non-negatif, sementara model NB terkadang salah mengklasifikasikannya karena keterbatasan dalam menangkap konteks implisit.

Secara keseluruhan, hasil analisis menunjukkan bahwa sentimen negatif umumnya terkait dengan gangguan teknis, kesulitan dalam proses verifikasi, dan lambatnya tindak lanjut aduan. Implikasi dari temuan ini mencakup perbaikan stabilitas aplikasi, penyederhanaan proses pendaftaran, serta peningkatan sistem respons dan pelatihan petugas pelayanan. Sebaliknya, temuan kata-kata non-negatif dapat dimanfaatkan untuk mengidentifikasi aspek pelayanan yang sudah berjalan baik, sehingga layak untuk dipertahankan dan dikembangkan lebih lanjut dalam pembaruan aplikasi.

4. Kesimpulan

Hasil penelitian menunjukkan bahwa analisis sentimen dengan pendekatan *machine learning* dapat memberikan wawasan yang lebih mendalam mengenai pengalaman pengguna terhadap pelayanan SP4N-LAPOR!. Dari ketiga skenario yang dilakukan, model NBSVM dengan fitur ekspansi *N-gram* (kombinasi *unigram* dan *bigram*) memberikan performa terbaik, dengan nilai *G-Mean* 0.8451, *Sensitivity* sebesar 0.8227 dan *F2 Score* sebesar 0.8215. Hal ini menunjukkan bahwa model NBSVM lebih andal dalam mengklasifikasikan sentimen pengguna dibandingkan jika masing-masing model (NB dan SVM) berdiri sendiri.

Selain itu, hasil analisis menunjukkan bahwa tantangan utama dalam penggunaan aplikasi masih berkaitan dengan aspek teknis, seperti kendala pada server, proses verifikasi, serta lambatnya respons terhadap pengaduan. Secara teknis, temuan ini memberikan kontribusi praktis yang penting bagi pengelola SP4N-LAPOR!, antara lain melalui

pemanfaatan hasil klasifikasi sentimen sebagai sistem peringatan dini (*early warning system*) untuk mengidentifikasi dan memprioritaskan aduan yang bernuansa negatif. Pada sisi pelayanan, klasifikasi sentimen juga dapat berfungsi sebagai alat evaluasi bagi penyelenggara dalam upaya peningkatan kualitas respons terhadap pengaduan masyarakat. Dengan demikian, analisis sentimen tidak hanya memberikan gambaran menyeluruh mengenai persepsi publik terhadap SP4N-LAPOR!, tetapi juga dapat digunakan sebagai dasar pengambilan keputusan berbasis data dalam pengelolaan pengaduan pelayanan publik digital. Harapannya, hasil ini dapat dimanfaatkan oleh pengelola aplikasi untuk merumuskan langkah-langkah peningkatan kualitas pelayanan secara berkelanjutan. Di samping itu, hasil penelitian ini juga dapat memperkuat peran Ombudsman RI sebagai lembaga pengawas pelaksanaan SP4N-LAPOR!, khususnya dalam memastikan bahwa setiap pengaduan masyarakat ditindaklanjuti secara cepat, akurat, dan akuntabel sesuai dengan prinsip pelayanan publik yang baik.

Namun demikian, perlu dicatat bahwa sumber data penelitian ini hanya terbatas dari aplikasi *mobile* SP4N-LAPOR! berbasis Android. Oleh karena itu, saran untuk penelitian mendatang dapat mempertimbangkan tambahan data dari aplikasi web. Apabila jumlah data yang digunakan nantinya semakin banyak, maka model klasifikasi juga dapat dikembangkan dengan teknik yang lebih kompleks, seperti *deep learning*. Hal ini tentunya akan meningkatkan akurasi analisis sentimen. Selain itu, penelitian selanjutnya juga dapat melakukan analisis sentimen berbasis aspek (*aspect-based sentiment analysis*), untuk memahami lebih dalam faktor-faktor spesifik yang mempengaruhi persepsi pengguna terhadap pelayanan. Dengan pendekatan yang lebih holistik, penelitian di masa depan dapat semakin memperkuat peran teknologi dalam mendukung transparansi dan akuntabilitas pelayanan publik di Indonesia.

Daftar Pustaka:

- Abbas, M., Ali, K., Jamali, A., Ali Memon, K., & Aleem Jamali, A. (2019). Multinomial Naive Bayes Classification Model for Sentiment Analysis. *IJCSNS International Journal of Computer Science and Network Security*, 19(3), 62. <https://doi.org/10.13140/RG.2.2.30021.40169>
- Adyatma, A. D., Afuan, L., & Maryanto, E. (2023). The Effect of Unigram and Bigram in The Naïve Bayes Multinomial for Analyzing of Comment Sentiment of Gojek Application in Google Play Store. *Jurnal Teknik Informatika (Jutif)*, 4(6), 1535–1540. <https://doi.org/10.52436/1.jutif.2023.4.6.1310>
- A'la, F. Y. (2022). Indonesian Sentiment Analysis towards MyPertamina Application Reviews by Utilizing Machine Learning Algorithms. *Journal of Informatics Information System Software Engineering and Applications (INISTA)*, 5(1), 80–91. <https://doi.org/10.20895/inista.v5i1.838>
- Alqaryouti, O., Siyam, N., Abdel Monem, A., & Shaalan, K. (2024). Aspect-based sentiment analysis using smart government review data. *Applied Computing and Informatics*, 20(1–2), 142–161. <https://doi.org/10.1016/j.aci.2019.11.003>
- Anreaja, L. J., Harefa, N. N., Galih, J., Negara, P., Nathan, V., Pribyantara, H., & Prasetyo, A. B. (2022). *JISA (Jurnal Informatika dan Sains) Naive Bayes and Support Vector Machine Algorithm for Sentiment Analysis Opensea Mobile Application Users in Indonesia*.
- Bekkar, M., Djema, H., Kheliouane Djemaa, D., & Akrouf Alitouche, D. (2013). *Evaluation measures for models assessment over imbalanced data sets*. 3(10). www.iiste.org
- Das, M., Kamalanathan, S., & Alphonse, P. (2020). *A Comparative Study on TF-IDF feature Weighting Method and its Analysis using Unstructured Dataset*.
- Faisal, M. R. (2019). *Belajar Data Science: Klasifikasi dengan Bahasa Pemrograman R*. <https://www.researchgate.net/publication/312160783>
- Hakim, S. N., Putra, A. J., & Khasanah, A. U. (2021). Sentiment analysis on myindihome user reviews using support vector machine and naïve bayes classifier method. *International Journal of Industrial Optimization*, 2(2), 151–164.
- Herlinawati, N., Yuliani, Y., Faizah, S., Gata, W., Komputer STMIK Nusa Mandiri Jl Damai No, I., Jati Barat, W., & Selatan, J. (2020). *Analisis Sentimen Zoom Cloud Meetings Di Play Store Menggunakan Naïve Bayes Dan Support Vector Machine* (Vol. 5, Issue 2).
- Hermanto, Kuntoro, A. Y., Asra, T., Pratama, E. B., Effendi, L., & Ocanitra, R. (2020). Gojek and Grab User Sentiment Analysis on Google Play Using Naive Bayes Algorithm and Support Vector Machine Based Smote Technique. *Journal of Physics: Conference Series*, 1641(1). <https://doi.org/10.1088/1742-6596/1641/1/012102>
- Lestari, S. D., & Setiawan, E. B. (2022). Sentiment Analysis Based on Aspects Using FastText Feature Expansion and NBSVM Classification Method. *Journal of Computer System and Informatics (JoSYC)*, 3(4), 469–477. <https://doi.org/10.47065/josyc.v3i4.2202>
- Muhammad, A. N. (2019). Sentiment Analysis of Positive and Negative of YouTube Comments Using Naive Bayes-Support Vector Machine (NBSVM) Classifier. *ICOMITEE 2019*.
- Mustakim, H., & Priyanta, S. (2022). Aspect-Based Sentiment Analysis of KAI Access Reviews Using NBC and SVM. *IJCCS (Indonesian*

- Journal of Computing and Cybernetics Systems*, 16(2), 113.
<https://doi.org/10.22146/ijccs.68903>
- Nawaz, S. (2023). A Comparative Study of Machine Learning Algorithms for Sentiment Analysis. *International Journal of Research Publication and Reviews Journal Homepage: Www.Ijrpr.Com*, 4(8), 1999–2003.
www.ijrpr.com
- Nurandi. (2015, January 27). *Membuat Word Cloud dengan R*.
- Palomino, M. A., & Aider, F. (2022). Evaluating the Effectiveness of Text Pre-Processing in Sentiment Analysis. *Applied Sciences (Switzerland)*, 12(17).
<https://doi.org/10.3390/app12178765>
- Permen PAN & RB No. 46 Tentang Road Map Sistem Pengelolaan Pengaduan Pelayanan Publik Nasional Tahun 2020-2024, Pub. L. No. 46 (2020).
<https://peraturan.bpk.go.id/Details/143742/permen-pan-rb-no-46-tahun-2020>
- Prabowo, R., & Thelwall, M. (2009). Sentiment analysis: A combined approach. *Journal of Informetrics*, 3(2), 143–157.
- Prasetyo, B., Alamsyah, Muslim, M. A., & Baroroh, N. (2021). Evaluation performance recall and F2 score of credit card fraud detection unbalanced dataset using SMOTE oversampling technique. *Journal of Physics: Conference Series*, 1918(4).
<https://doi.org/10.1088/1742-6596/1918/4/042002>
- Premaiswari, W., & Made, N. (2024). *The Effectiveness of the SP4N-LAPOR! as a National Public Service Com-plaint Management Application*. 28(1), 36–51.
<https://journal.ugm.ac.id/jkap>
- Pusvita, V., & Muttaqin, M. (2023). Satu Dekade LAPOR: Tren Penggunaan Kanal Layanan SP4N. *Jurnal Pekommas*, 8(1), 1–10.
<https://doi.org/10.56873/jpkm.v8i1.5138>
- Saputro, H. T., & Hermawan, A. (2021). The Accuracy Improvement of Text Mining Classification on Hospital Review through The Alteration in The Preprocessing Stage. In *International Journal of Computer and Information Technology* (Vol. 10, Issue 4).
www.ijcit.com140
- Truong, B., Zheng, J., Hornsby, L., Fox, B., Chou, C., & Qian, J. (2024). Development and Validation of Machine Learning Algorithms to Predict 1-Year Ischemic Stroke and Bleeding Events in Patients with Atrial Fibrillation and Cancer. *Cardiovascular Toxicology*, 24(4), 365–374.
<https://doi.org/10.1007/s12012-024-09843-8>
- Wang, S., Dai, Y., Shen, J., & Xuan, J. (2021). Research on expansion and classification of imbalanced data based on SMOTE algorithm. *Scientific Reports*, 11(1).
<https://doi.org/10.1038/s41598-021-03430-5>
- Wang, S., & Manning, C. D. (2012). *Baselines and Bigrams: Simple, Good Sentiment and Topic Classification*. <http://www.cs.uic.edu/>
- Wildhani, A. M., Nurmandi, A., Misran, & Subekti, D. (2023). Application of SP4N-LAPOR to Improve the Quality of Public Services and Information (Case Study in Indonesia). *Communications in Computer and Information Science*, 1835 CCIS, 297–305.
https://doi.org/10.1007/978-3-031-36001-5_38
- Xu, W., Chen, J., Ding, Z., & Wang, J. (2024). *Text Sentiment Analysis and Classification Based on Bidirectional Gated Recurrent Units (GRUs) Model*.

Halaman ini sengaja dikosongkan