NOVAKomputa
Jurnal Ilmu Komputer dan Informatika

E-ISSN: 0000-0000
Volume 1 Nomor 2 | Juni 2026

Available online: https://jurnal.gsminnovation.com/index.php/novakomputa/article/view/9

Klasifikasi Pesan Pengaduan Masyarakat Berbasis Telegram Bot Menggunakan TF-IDF dan Support Vector Machine

Shabira[1]*, Imam Muslem[2], Riyadhul Fajri[3]

[1] Program Studi Informatika, Universitas Almuslim, Jl. Tengku Abdurrahman No. 37, Matangglumpangdua, Peusangan, Bireuen, Aceh 24261, Indonesia. e-mail: shabira.net2019@gmail.com

[2] Program Studi Informatika, Universitas Almuslim, Jl. Tengku Abdurrahman No. 37, Matangglumpangdua, Peusangan, Bireuen, Aceh 24261, Indonesia. e-mail: imamtkj@gmail.com

[3] Program Studi Informatika, Universitas Almuslim, Jl. Tengku Abdurrahman No. 37, Matangglumpangdua, Peusangan, Bireuen, Aceh 24261, Indonesia. e-mail: fajri071113@gmail.com

*Corresponding Author:
Shabira
shabira.net2019@gmail.com
ORCID: https://orcid.org/0009-0004-6837-3978

ARTICLE INFO

Article History:
Received : 29 April 2026
Revised : 12 Mei 2026
Accepted : 11 Juni 2026

Keywords:
Public Complaint Classification, Telegram Bot, Text Mining, TF-IDF, Support Vector Machine

Kata kunci:
Klasifikasi Pengaduan Masyarakat, Telegram Bot, Text Mining, TF-IDF, Support Vector Machine

NOVAKOMPUTA
Jurnal Ilmu Komputer dan Informatika
Vol 01 No 02 | Juni 2026

ABSTRACT

[Telegram Bot-Based Public Complaint Message Classification Using TF-IDF and Support Vector Machine] Public complaint management requires an efficient mechanism to classify and distribute reports to relevant agencies. Manual complaint handling is often prone to delays and misclassification, particularly when incoming messages vary in structure and content. This study aims to develop a Telegram Bot-based public complaint classification system using text mining, Term Frequency-Inverse Document Frequency (TF-IDF), and the Support Vector Machine (SVM) algorithm. The dataset used in this study consisted of 1,000 simulated complaint messages categorized into five agency classes, namely PUPR, DLHK, DINKES, DINSOS, and DISDIKBUD. The research stages included text preprocessing, feature extraction, model training, model evaluation, and Telegram Bot implementation. Text preprocessing was performed through cleaning, case folding, tokenization, stopword removal, and stemming using the Sastrawi library. The processed text was transformed into numerical features using TF-IDF with unigram and bigram representations. The dataset was divided into 80% training data and 20% testing data, and the classification model was built using SVM. The evaluation results showed that the model achieved a training accuracy of 99.7%, testing accuracy of 98.5%, precision of 99%, recall of 98%, and F1-score of 98%. The Telegram Bot implementation showed that the system could receive complaint messages, classify them automatically, and forward them to the relevant agency groups. These results indicate that the proposed system can improve the efficiency and accuracy of public complaint distribution.

ABSTRAK

Pengelolaan pengaduan masyarakat memerlukan mekanisme yang efisien untuk mengklasifikasikan dan mendistribusikan laporan kepada instansi yang relevan. Penanganan pengaduan secara manual sering menimbulkan keterlambatan dan kesalahan klasifikasi, terutama ketika pesan yang masuk memiliki struktur dan isi yang beragam. Penelitian ini bertujuan untuk mengembangkan sistem klasifikasi pengaduan masyarakat berbasis Telegram Bot menggunakan text mining, Term Frequency-Inverse Document Frequency (TF-IDF), dan algoritma Support Vector Machine (SVM). Dataset yang

digunakan dalam penelitian ini terdiri dari 1.000 pesan pengaduan simulasi yang dikelompokkan ke dalam lima kelas instansi, yaitu PUPR, DLHK, DINKES, DINSOS, dan DISDIKBUD. Tahapan penelitian meliputi prapemrosesan teks, ekstraksi fitur, pelatihan model, evaluasi model, dan implementasi Telegram Bot. Prapemrosesan teks dilakukan melalui cleaning, case folding, tokenization, stopword removal, dan stemming menggunakan pustaka Sastrawi. Teks hasil prapemrosesan kemudian diubah menjadi fitur numerik menggunakan TF-IDF dengan representasi unigram dan bigram. Dataset dibagi menjadi 80% data latih dan 20% data uji, sedangkan model klasifikasi dibangun menggunakan SVM. Hasil evaluasi menunjukkan bahwa model memperoleh training accuracy sebesar 99,7%, testing accuracy sebesar 98,5%, precision sebesar 99%, recall sebesar 98%, dan F1-score sebesar 98%. Implementasi Telegram Bot menunjukkan bahwa sistem mampu menerima pesan pengaduan, melakukan klasifikasi secara otomatis, dan meneruskan laporan ke grup instansi yang sesuai. Hasil ini menunjukkan bahwa sistem yang diusulkan dapat meningkatkan efisiensi dan akurasi distribusi pengaduan masyarakat.

PENDAHULUAN

Pengaduan masyarakat merupakan salah satu bentuk partisipasi publik dalam menyampaikan permasalahan di lingkungan sekitar [1], [2]. Namun, pengelolaan pengaduan yang masih dilakukan secara manual sering menimbulkan kendala seperti keterlambatan, kesalahan pengelompokan, dan kurang efisiennya distribusi laporan ke instansi terkait [3], [4]. Oleh karena itu, diperlukan sistem yang mampu mengklasifikasikan pengaduan secara otomatis agar proses penanganan menjadi lebih cepat dan tepat [5]. Dalam hal ini, pendekatan *text mining* menjadi solusi yang efektif untuk mengolah data teks tidak terstruktur [6], dengan salah satu metode yang banyak digunakan adalah Support Vector Machine (SVM) karena kemampuannya dalam menangani data berdimensi tinggi [7], [8], [9], [10].

Beberapa penelitian sebelumnya telah membuktikan efektivitas SVM dalam klasifikasi teks dan berhasil mengidentifikasi *cyber hate speech* dengan akurasi 96% menggunakan kombinasi TF-IDF dan SVM [9], [10], [11], [12]. Penelitian lain juga menunjukkan bahwa SVM mampu mengklasifikasikan pengaduan publik dengan baik setelah melalui tahapan preprocessing [13]. Mengimplementasikan chatbot Telegram berbasis SVM dengan akurasi 98,92%, yang menunjukkan potensi integrasi antara klasifikasi teks dan sistem otomatis [14]. Penelitian lain juga membuktikan bahwa SVM memiliki performa lebih baik dibandingkan metode lain seperti Naïve Bayes dalam beberapa kasus klasifikasi teks [15], [16]. Sementara itu, penggunaan TF-IDF juga memberikan hasil yang lebih konsisten dibandingkan Word2Vec dalam klasifikasi teks**,** sehingga mendukung pemilihan metode pada penelitian ini [17].

Berbagai penelitian lain juga memperkuat pentingnya pemilihan metode dan fitur dalam klasifikasi teks[18]. Menerapkan SVM untuk analisis sentimen dengan TF-IDF dan memperoleh hasil yang cukup baik meskipun terdapat tantangan ketidakseimbangan data [19]. SVM bisa disebut juga lebih unggul dibandingkan metode lain dalam analisis sentimen pengguna aplikasi [20]. Beberapa penelitian lain juga menunjukkan penerapan SVM dalam berbagai bidang analisis teks dan klasifikasi data. Penggunaan SVM yang dikombinasikan dengan TF-IDF, SMOTE, dan Particle Swarm Optimization untuk menganalisis sentimen politik di media sosial, mampu memberikan hasil akurasi mencapai 87,50% [21]. Sementara itu, menerapkan SVM dalam analisis sentimen terhadap layanan uang elektronik OVO dan memperoleh akurasi tertinggi sebesar 98,7%, yang menunjukkan efektivitas SVM dalam klasifikasi data teks dari media sosial [22].

Pemanfaatan metode SVM dengan pembobotan TF-IDF untuk menganalisis sentimen opini alumni, dengan hasil akurasi sebesar 87,3%, yang menunjukkan bahwa SVM mampu memberikan performa yang stabil dalam analisis teks [23]. Tidak hanya pada data teks, SVM juga diterapkan pada bidang lain seperti kesehatan, yaitu mereka menggunakan SVM untuk klasifikasi diagnosis kanker payudara. Hasil penelitian tersebut menunjukkan bahwa SVM mampu mengidentifikasi pola data dan memberikan prediksi yang akurat, sehingga memperkuat bahwa algoritma ini memiliki fleksibilitas dan keandalan dalam berbagai jenis permasalahan klasifikasi [24].

Berdasarkan permasalahan dan studi-studi sebelumnya yang telah dibahas, penelitian ini bertujuan untuk merancang sistem klasifikasi keluhan masyarakat menggunakan metode TF-IDF dan

algoritma Support Vector Machine (SVM), yang terintegrasi ke dalam bot Telegram. Sistem yang dikembangkan diharapkan dapat secara otomatis mengkategorikan keluhan berdasarkan instansi pemerintah yang bersangkutan. Ruang lingkup penelitian ini mencakup tahap prapemrosesan, penentuan bobot kata menggunakan TF-IDF, proses klasifikasi dengan SVM, serta implementasi sistem pada platform Telegram. Kontribusi utama yang ditawarkan dalam studi ini adalah realisasi sistem klasifikasi yang terintegrasi dengan media komunikasi, sehingga mendukung peningkatan efisiensi dan mempercepat penanganan keluhan masyarakat secara real-time.

A. Research Gap

Berdasarkan kajian terhadap penelitian-penelitian sebelumnya, terlihat bahwa metode Support Vector Machine (SVM) telah banyak digunakan dalam klasifikasi teks dan menunjukkan performa yang cukup baik dalam berbagai kasus, seperti klasifikasi cyber hate speech, pengaduan publik, serta analisis sentimen pada media sosial [11], [13], [16], [22]. Selain itu, kombinasi antara teknik preprocessing dan pembobotan TF-IDF juga terbukti mampu meningkatkan akurasi dalam proses klasifikasi teks [19]. Beberapa penelitian bahkan telah mengintegrasikan metode SVM dalam sistem berbasis chatbot untuk memahami intent pengguna secara otomatis [14].

Meskipun demikian, sebagian besar penelitian yang ada hingga saat ini masih berfokus terutama pada aspek klasifikasi model itu sendiri, dan hanya sedikit yang telah mengintegrasikan sistem klasifikasi tersebut secara langsung dengan platform komunikasi real-time yang digunakan oleh masyarakat. Selain itu, masih terdapat keterbatasan dalam mengakomodasi variasi masukan teks yang berada di luar kategori yang telah ditentukan sebelumnya, sehingga model cenderung memaksakan hasil klasifikasi ke salah satu kelas yang tersedia. Situasi ini menunjukkan adanya celah penelitian dalam pengembangan sistem klasifikasi keluhan yang tidak hanya unggul dalam hal akurasi model, tetapi juga terhubung langsung dengan media komunikasi dan mampu menangani masukan yang lebih dinamis.

B. Kebaruan dan Kontribusi Ilmiah

Penelitian ini memperkenalkan sebuah inovasi berupa sistem klasifikasi pengaduan publik yang menggabungkan metode TF-IDF dan algoritma Support Vector Machine (SVM) dengan platform bot Telegram sebagai sarana interaksi pengguna. Berbeda dengan penelitian sebelumnya yang umumnya hanya berfokus pada perancangan model klasifikasi, penelitian ini menerapkan model tersebut ke dalam sistem yang dapat langsung dimanfaatkan oleh pengguna untuk mengajukan pengaduan dan menerima tanggapan otomatis.

Selain itu, kontribusi ilmiah yang ditawarkan dalam studi ini terletak pada penerapan pendekatan end-to-end, yang mencakup semua tahap mulai dari prapemrosesan data teks dan penentuan bobot kata menggunakan TF-IDF hingga proses klasifikasi menggunakan SVM, yang semuanya terintegrasi ke dalam sistem berbasis komunikasi real-time. Studi ini juga menyajikan gambaran kinerja model saat diterapkan dalam kondisi dunia nyata, termasuk kemungkinan terjadinya kesalahan klasifikasi akibat keragaman pola teks yang masuk. Dengan demikian, penelitian ini diharapkan dapat berkontribusi pada pengembangan sistem pemrosesan pengaduan publik yang lebih efisien, otomatis, dan praktis dalam ranah pelayanan publik.

METODE

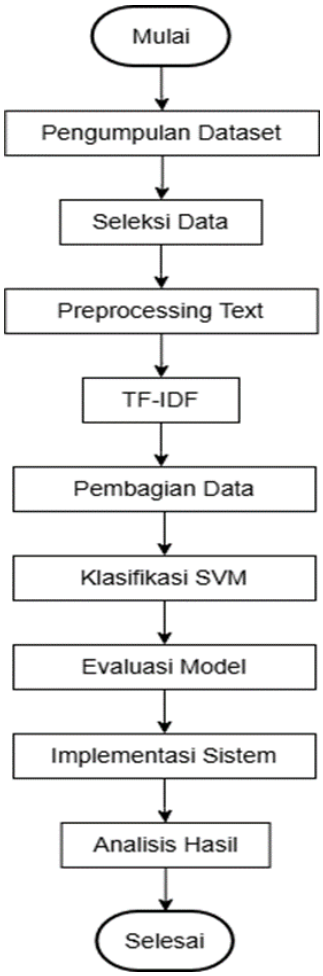
Penelitian ini menggunakan pendekatan kuantitatif komputasional dengan metode penambangan teks untuk mengklasifikasikan keluhan publik berbasis teks. Penelitian ini dikategorikan sebagai penelitian eksperimental, karena melibatkan fase pelatihan dan pengujian model menggunakan algoritma pembelajaran mesin. Pendekatan ini dipilih karena penelitian ini berfokus pada pemrosesan data teks menjadi informasi yang dapat dijadikan dasar untuk pengambilan keputusan otomatis, khususnya klasifikasi keluhan ke dalam kategori instansi pemerintah yang sesuai.

Dalam implementasinya, penelitian ini menggunakan bahasa pemrograman Python, yang dijalankan dalam lingkungan pengembangan Jupyter Notebook. Python dipilih karena keunggulannya dalam menyediakan berbagai pustaka pendukung untuk pemrosesan data dan pembelajaran mesin. Beberapa pustaka yang digunakan antara lain pandas untuk pemrosesan data, NumPy untuk komputasi numerik, dan scikit-learn untuk mengimplementasikan algoritma TF-IDF dan Support Vector Machine (SVM). Selain itu, sistem ini terintegrasi dengan Telegram Bot API, yang memungkinkan interaksi

langsung antara pengguna dan sistem. Perangkat keras yang digunakan dalam penelitian ini terdiri dari komputer atau laptop dengan spesifikasi yang memadai untuk menangani komputasi data teks dan pelatihan model.

A. Tahapan Penelitian

Penelitian ini dilakukan melalui serangkaian tahap yang dirancang secara sistematis untuk memastikan bahwa seluruh proses berjalan secara terstruktur sekaligus menghasilkan model dengan kinerja optimal. Tahap awal dimulai dengan pengumpulan dan pengorganisasian kumpulan data keluhan masyarakat. Selanjutnya, data yang terkumpul diproses pada tahap preprocessing untuk membersihkan teks dari unsur-unsur yang dianggap tidak relevan. Setelah proses ini, dilakukan pembobotan kata menggunakan metode TF-IDF untuk mengubah data teks menjadi bentuk numerik. Data numerik yang dihasilkan kemudian digunakan sebagai input untuk proses pelatihan model, dengan memanfaatkan algoritma Support Vector Machine (SVM). Model yang telah dilatih kemudian diuji menggunakan data uji untuk mengukur akurasi dan kinerja klasifikasinya. Tahap akhir dari rangkaian penelitian ini adalah implementasi model ke dalam sistem berbasis bot Telegram sehingga dapat diakses dan digunakan langsung oleh pengguna.



Gambar 1. Alur Penelitian

B. Dataset Penelitian

Dataset yang digunakan dalam penelitian ini merupakan data pengaduan masyarakat yang disusun secara mandiri (data simulasi). Penyusunan dataset dilakukan dengan mempertimbangkan karakteristik pengaduan yang umum disampaikan oleh masyarakat, seperti permasalahan infrastruktur, lingkungan, sosial, dan layanan publik lainnya. Penggunaan data simulasi dilakukan karena keterbatasan ketersediaan dataset publik yang sesuai dengan kebutuhan penelitian.

Dataset terdiri dari dua atribut utama, yaitu teks hasil preprocessing (*processed_text*) dan label kategori (*label*). Label yang digunakan dalam penelitian ini mencerminkan kategori dinas yang

menangani pengaduan, seperti Dinas Pekerjaan Umum (PUPR), Dinas Lingkungan Hidup (DLHK), Dinas Sosial (DINSOS), Dinas Kesehatan (DINKES), dan Dinas Pendidikan (DISDIKBUD).
Jumlah dataset yang digunakan dalam penelitian ini adalah **1000** data. Dataset tersebut kemudian dibagi menjadi dua bagian, yaitu data latih (*training data*) sebesar **80%** dan data uji (*testing data*) sebesar **20%**. Pembagian ini bertujuan untuk memastikan bahwa model dapat dilatih dengan baik serta diuji menggunakan data yang belum pernah dilihat sebelumnya, sehingga hasil evaluasi lebih objektif.

Tabel 1. Pembagian Dataset

Kelas	Jumlah Dataset	Data Latih	Data Uji
PUPR	200	160	40
DLHK	200	160	40
DINKES	200	160	40
DINSOS	200	160	40
DISDIKBUD	200	160	40
Total	**1000**	**800**	**200**

C. Preprocessing Data

Preprocessing merupakan tahap penting dalam text mining yang bertujuan untuk membersihkan dan menyiapkan data teks sebelum dilakukan proses analisis lebih lanjut. Data teks yang berasal dari pengaduan masyarakat umumnya masih mengandung berbagai noise, seperti tanda baca, angka, serta kata-kata yang tidak relevan, sehingga perlu dilakukan pembersihan. Hasil dari tahapan preprocessing ini berupa teks yang lebih bersih, terstruktur, dan siap digunakan dalam proses pembobotan fitur. Beberapa langkah yang dilakukan dalam proses ini dapat dilihat dalam Tabel 2.

Tabel 2. Langkah-langkah Preprocessing

Tahapan	Deskripsi Singkat	Contoh Teks
Data Awal	Teks pengaduan asli dari pengguna.	Warga melaporkan! Jalan rusak parah di daerah ini.
Cleaning	Menghapus tanda baca seperti titik, koma atau tanda seru, simbol, dan karakter tidak penting menggunakan *regular expression* dari modul *re.*	Warga melaporkan Jalan rusak parah di daerah ini
Case Folding	Mengubah semua huruf menjadi huruf kecil menggunakan fungsi bawaan Python yaitu *.lower().*	warga melaporkan jalan rusak parah di daerah ini
Tokenizing	Memecah kalimat menjadi kata-kata (token) menggunakan fungsi bawaan Python yaitu *split().*	['warga', 'melaporkan', 'jalan', 'rusak', 'parah', 'di', 'daerah', 'ini']
Stopword Removal	Menghapus kata umum yang tidak penting seperti "yang", "di", "dan" menggunakan kamus stopwords dari nltk.	warga, melaporkan, jalan, rusak, parah, daerah
Stemming	Mengubah kata ke bentuk dasar menggunakan library Sastrawi, khususnya *StemmerFactory.*	warga, lapor, jalan, rusak, parah, daerah

D. Pembobotan TF-IDF

Setelah data teks melalui tahap preprocessing, langkah selanjutnya adalah mengubah data tersebut ke dalam bentuk numerik menggunakan metode Term Frequency-Inverse Document Frequency (TF-IDF). Metode ini digunakan untuk menentukan tingkat kepentingan suatu kata dalam sebuah dokumen relatif terhadap seluruh dokumen dalam dataset.

Secara matematis, rumus perhitungan TF-IDF dinyatakan sebagai berikut:

$$TF - IDF(t,d) = TF\ (t,d)\ x\ \left(\frac{N}{df(t)}\right) \quad (1)$$

Dimana:
a. *t* adalah term atau kata

b. d adalah dokumen
c. $tf(t,d)$ adalah jumlah kemunculan kata t dalam dokumen d
d. $\sum tf(t,d)$ d adalah total kata dalam dokumen
e. N adalah jumlah seluruh dokumen
f. $td(t,d)$ adalah jumlah dokumen yang mengandung kata t

Nilai TF-IDF akan tinggi jika suatu kata sering muncul dalam satu dokumen tetapi jarang muncul dalam dokumen lain. Sebaliknya, jika suatu kata sering muncul di banyak dokumen, maka nilainya akan kecil karena dianggap kurang informatif. Hasil dari proses ini adalah representasi vektor numerik untuk setiap dokumen, yang selanjutnya digunakan sebagai input dalam proses klasifikasi.

E. Klasifikasi Menggunakan Support Vector Machine (SVM)

Support Vector Machine (SVM) merupakan algoritma supervised learning yang digunakan untuk melakukan klasifikasi data. SVM bekerja dengan mencari hyperplane terbaik yang dapat memisahkan data ke dalam kelas-kelas yang berbeda dengan margin maksimum.

Persamaan dasar SVM dinyatakan sebagai berikut:

$$f(x) = w \,.\, x + b \qquad (2)$$

Dimana:
a. x adalah vektor fitur hasil TF-IDF
b. w adalah bobot (weight)
c. b adalah bias
d. $f(x)$ adalah fungsi keputusan

Tujuan dari SVM adalah menemukan nilai w dan b yang dapat memaksimalkan margin antar kelas. Data yang berada paling dekat dengan hyperplane disebut sebagai *support vector*, yang memiliki peran penting dalam menentukan batas pemisah. Dalam penelitian ini, proses klasifikasi dilakukan melalui dua tahap utama. Tahap pertama adalah *training*, yaitu proses pelatihan model menggunakan data latih untuk mempelajari pola dari masing-masing kelas. Tahap kedua adalah *testing*, yaitu proses pengujian model menggunakan data uji untuk mengetahui kemampuan model dalam mengklasifikasikan data baru.

F. Evaluasi Model

Evaluasi model dilakukan untuk mengukur kinerja sistem dalam melakukan klasifikasi. Metode evaluasi yang digunakan adalah confusion matrix, yang memberikan gambaran mengenai jumlah prediksi yang benar dan salah untuk setiap kelas.
Berdasarkan confusion matrix, dihitung beberapa metrik evaluasi, yaitu:
a. *Accuracy,* untuk mengukur tingkat ketepatan kseluruhan model
b. *Precission,* untuk mengukur ketepatan prediksi pada setiap kelas
c. *Recall,* untuk mengukur kemamapuan model dalam menemukan data yang relevan
d. *F1-Score,* sebagai keseimbangn antara precission dan recall

1) Accuracy $$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

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

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

4) F1-Score $$F1 = 2\ x\ \frac{Precision\ x\ Recall}{Precision + Recall}$$

Dimana:
a. TP (True Positive): data yang diprediksi benar
b. TN (True Negative): data yang diprediksi benar sebagai bukan kelas
c. FP (False Positive): data yang salah diprediksi
d. FN (False Negative): data yang tidak terdeteksi

HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil yang diperoleh dari klasifikasi keluhan masyarakat menggunakan metode TF-IDF dan algoritma Support Vector Machine (SVM), beserta pembahasan mengenai kinerja model yang dihasilkan. Temuan penelitian ini dianalisis berdasarkan beberapa aspek, yaitu pembagian data, proses klasifikasi, dan evaluasi menggunakan metrik akurasi, presisi, recall, dan F1-score. Selain itu, analisis ini berfokus pada matriks kebingungan untuk mengidentifikasi pola kesalahan klasifikasi yang muncul selama proses pengujian. Lebih lanjut, bagian ini merinci implementasi sistem berbasis bot Telegram sebagai aplikasi praktis dari model dalam kondisi dunia nyata. Sebagai pelengkap, dilakukan perbandingan dengan temuan penelitian sebelumnya untuk menyoroti posisi dan kontribusi yang ditawarkan oleh penelitian ini.

A. Struktur Dataset

Dataset yang digunakan dalam penelitian ini terdiri dari tiga komponen utama, yaitu nomor, deskripsi teks dan disposisi dinas sebagai label target. Kolom deskripsi menjadi bagian yang paling penting dalam proses klasifikasi karena berisi uraian permasalahan yang disampaikan oleh masyarakat. Sedangkan kolom label berfungsi sebagai penanda kelas yang menunjukkan instansi atau dinas yang bertanggung jawab dalam menangani laporan tersebut. Struktur dan contoh isi dari dataset ditunjukkan pada Tabel 3.

Tabel 3. Contoh isi dataset

No	Text	Label
1	Warga melaporkan masalah jalan yang berdampak pada pelayanan masyarakat di lingkungan sekitar.	PUPR
….	….	….
1000	Terdapat keluhan masyarakat mengenai pendidikan dan bantuan yang perlu segera ditindaklanjuti.	DISDIKBUD

Tabel 3 menyajikan contoh isi dataset yang digunakan dalam penelitian, yang terdiri atas tiga atribut utama, yaitu nomor data, teks pengaduan, dan label kategori instansi tujuan. Kolom Text berisi pesan pengaduan masyarakat dalam bentuk kalimat, misalnya laporan mengenai permasalahan jalan atau keluhan terkait pendidikan dan bantuan. Sementara itu, kolom Label menunjukkan kelas atau kategori instansi yang bertanggung jawab menangani pengaduan tersebut, seperti PUPR untuk laporan infrastruktur jalan dan DISDIKBUD untuk laporan yang berkaitan dengan pendidikan. Contoh data pada tabel ini menunjukkan bahwa setiap teks pengaduan telah dipasangkan dengan label yang sesuai, sehingga dapat digunakan sebagai data latih dan data uji dalam proses klasifikasi menggunakan metode TF-IDF dan algoritma Support Vector Machine.

B. Preprocessing

Hasil preprocessing menunjukkan bahwa data teks pengaduan yang awalnya tidak terstruktur berhasil dibersihkan dan dibuat lebih konsisten. Proses ini meliputi penghapusan tanda baca dan simbol menggunakan *regular expression*, perubahan huruf menjadi kecil (*case folding*), pemecahan kata (*tokenizing*), penghapusan kata tidak penting (*stopword removal*), serta pengubahan kata ke bentuk dasar menggunakan Sastrawi (*stemming*). Hasilnya adalah teks yang lebih sederhana dan siap untuk dianalisis lebih lanjut. Contoh hasil sebelum dan sesudah preprocessing ditunjukkan pada Tabel 4.

Tabel 4. Sebelum dan sesudah Preprocessing

No	Sebelum Preprocessing	Sesudah Preprocessing
1	Warga melaporkan masalah jalan yang berdampak pada pelayanan masyarakat di lingkungan sekitar.	warga lapor jalan dampak layan masyarakat lingkung
….	….	….
1000	Terdapat keluhan masyarakat mengenai pendidikan dan bantuan yang perlu segera ditindaklanjuti.	keluh masyarakat didik bantu ditindaklanjuti

Tabel 4 menunjukkan contoh perubahan teks pengaduan sebelum dan sesudah melalui proses preprocessing. Pada kolom Sebelum Preprocessing, teks masih berbentuk kalimat asli yang mengandung kata-kata umum dan struktur bahasa alami, seperti “Warga melaporkan masalah jalan yang berdampak

pada pelayanan masyarakat di lingkungan sekitar." Setelah dilakukan preprocessing, teks menjadi lebih ringkas dan terstandar, misalnya berubah menjadi "warga lapor jalan dampak layan masyarakat lingkung". Perubahan ini menunjukkan bahwa tahapan preprocessing telah menghilangkan kata-kata yang kurang relevan, menyeragamkan bentuk huruf, serta mengubah kata ke bentuk dasar. Dengan demikian, hasil pada Tabel 4 memperlihatkan bahwa data teks telah disederhanakan agar lebih siap digunakan dalam proses ekstraksi fitur TF-IDF dan klasifikasi menggunakan algoritma Support Vector Machine.

C. Ekstraksi Fitur TF-IDF

Setelah melalui tahap preprocessing, data teks kemudian masuk ke tahap ekstraksi fitur menggunakan metode TF-IDF untuk mengubah teks menjadi representasi numerik. Pada tahap ini digunakan pendekatan *kata tunggal (unigram)* seperti "warga" atau "lapor", dan gabungan dua kata yang berurutan *(bigram)* seperti "warga lapor" atau "jalan rusak". Penggunaan unigram dan bigram bertujuan untuk menangkap informasi tidak hanya dari kata tunggal, tetapi juga dari konteks antar kata. Hasil representasi ditampilkan pada Tabel 5.

Tabel 5. Hasil representasi TF-IDF

No.	aktivitas	aktivitas aspal	aktivitas bantu	aktivitas bau	…..	warga lapor	warga layan	warga lingkung	warga lokasi
1	0	0	0	0	…..	0.195	0	0	0
2	0	0	0	0	…..	0	0	0	0
3	0	0	0	0	…..	0.195	0	0	0
.....	…..	…..	…..	…..	…..	…..	…..	…..	…..
998	0	0	0	0	…..	0	0	0	0
999	0	0	0	0	…..	0	0	0	0
1000	0	0	0	0	…..	0	0	0	0

Nilai pada tabel 5 menunjukkan tingkat kepentingan suatu kata dalam dokumen tertentu; jika bernilai 0 berarti kata tersebut tidak muncul dalam dokumen, sedangkan nilai seperti 0.195 menunjukkan bahwa kata tersebut muncul dan memiliki tingkat kepentingan tertentu. Semakin besar nilainya, semakin penting kata tersebut dalam merepresentasikan isi dokumen. Tabel ini kemudian digunakan sebagai input numerik untuk proses klasifikasi menggunakan algoritma SVM.

D. Klasifikasi Menggunakan Support Vector Machine

Proses klasifikasi dalam penelitian ini dilakukan dengan menggunakan algoritma Support Vector Machine (SVM), sebuah metode pembelajaran terawasi yang umum diterapkan dalam bidang klasifikasi teks. Algoritma SVM bekerja dengan memanfaatkan kumpulan data berlabel untuk membangun model yang mampu mengenali dan membedakan setiap kategori berdasarkan pola yang telah dipelajarinya. Dalam penelitian ini, input yang digunakan terdiri dari hasil yang diberi bobot TF-IDF, sehingga setiap dokumen keluhan disajikan dalam bentuk vektor numerik yang mewakili bobot kata-kata penting yang terkandung dalam dokumen tersebut.

Secara konseptual, SVM bekerja dengan menentukan garis batas yang dikenal sebagai hyperplane untuk memisahkan data ke dalam kelas-kelas tertentu. Dalam konteks klasifikasi teks yang melibatkan jumlah fitur (kata) yang sangat besar, hiperbidang tersebut bukan lagi sekadar garis, melainkan sebuah bidang dalam ruang berdimensi tinggi. SVM memilih hiperbidang yang paling optimal-yaitu, hiperbidang dengan margin terbesar relatif terhadap titik-titik data dari setiap kelas. Semakin lebar marginnya, semakin jelas pemisahan antar kelas, sehingga menghasilkan model dengan kemampuan generalisasi yang lebih baik ketika dihadapkan pada data baru.

Dalam proses pelatihan (*training*), SVM akan menganalisis data latih untuk menemukan parameter terbaik, yaitu bobot (*weight*) dan bias yang digunakan dalam fungsi keputusan. Fungsi keputusan tersebut dinyatakan sebagai $f(x) = w \,.\, x + b$, di mana x merupakan vektor fitur hasil TF-IDF, w adalah bobot yang menunjukkan kontribusi masing-masing fitur, dan b adalah bias. Nilai dari fungsi ini digunakan untuk menentukan posisi suatu data terhadap hyperplane, sehingga dapat diketahui apakah data tersebut termasuk ke dalam kelas tertentu atau tidak.

Selain itu, dalam kasus klasifikasi multi-kelas seperti pada penelitian ini, SVM menerapkan strategi tertentu, seperti *one-vs-rest*, untuk membedakan lebih dari dua kategori. Setiap kelas akan dibandingkan dengan kelas lainnya untuk membentuk beberapa model klasifikasi, kemudian hasil akhir ditentukan berdasarkan model dengan nilai keputusan tertinggi.

Setelah proses pelatihan selesai, model kemudian diuji menggunakan data uji (*testing*). Pada tahap ini, setiap data uji akan diproses melalui tahapan yang sama, yaitu preprocessing dan pembobotan TF-IDF, kemudian dimasukkan ke dalam model SVM untuk diprediksi kelasnya. Hasil prediksi ini kemudian dibandingkan dengan label sebenarnya untuk mengevaluasi kinerja model. Dari proses ini dapat diketahui sejauh mana model mampu mengenali pola dalam data dan mengklasifikasikan pengaduan masyarakat dengan tepat.

E. Hasil Evaluasi

Evaluasi performa model pada penelitian ini diperoleh dari pengujian model SVM menggunakan data uji. Evaluasi dilakukan menggunakan metrik seperti accuracy, precision, recall, dan F1-score yang ditampilkan dalam bentuk *classification report*.

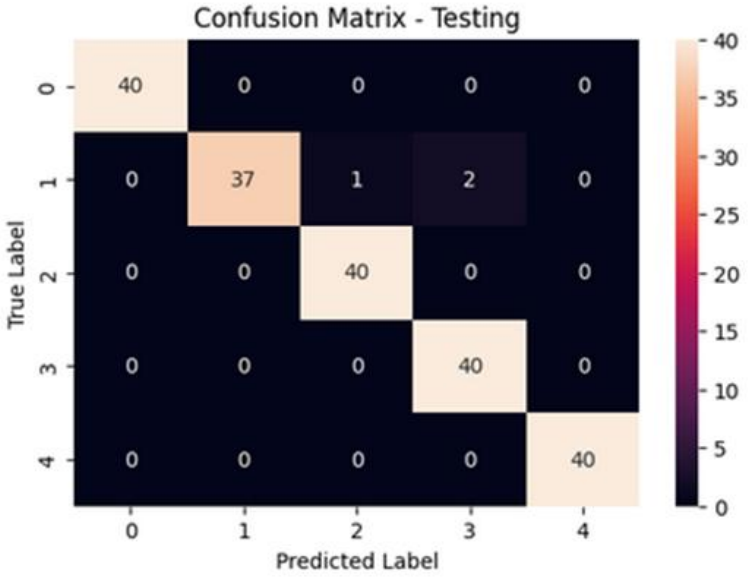
Tabel 6. Hasil evaluasi

Metrik	Evaluasi
Training Accuracy	0,997
Testing Accuracy	0.985
Precision	0.99
Recall	0.98
F1-score	0.98

Berdasarkan hasil pengujian yang ditunjukkan dalam tabel 6, model SVM menunjukkan performa yang sangat baik dengan training accuracy sebesar 0.997 dan testing accuracy sebesar 0.985. Selisih yang kecil antara training dan testing accuracy menjadi indikator bahwa model tidak mengalami overfitting. Secara keseluruhan, nilai accuracy mencapai 0.98, dengan macro average dan weighted average masing-masing sebesar 0.99 (precision), 0.98 (recall), dan 0.98 (F1-score), yang menandakan bahwa model memiliki kinerja yang stabil dalam mengklasifikasikan data pengaduan.

F. Analisis Confusion Matrix

Confusion matrix digunakan untuk menganalisis lebih detail hasil klasifikasi yang dilakukan oleh model. Matriks ini menunjukkan jumlah data yang diprediksi benar dan salah untuk setiap kelas.



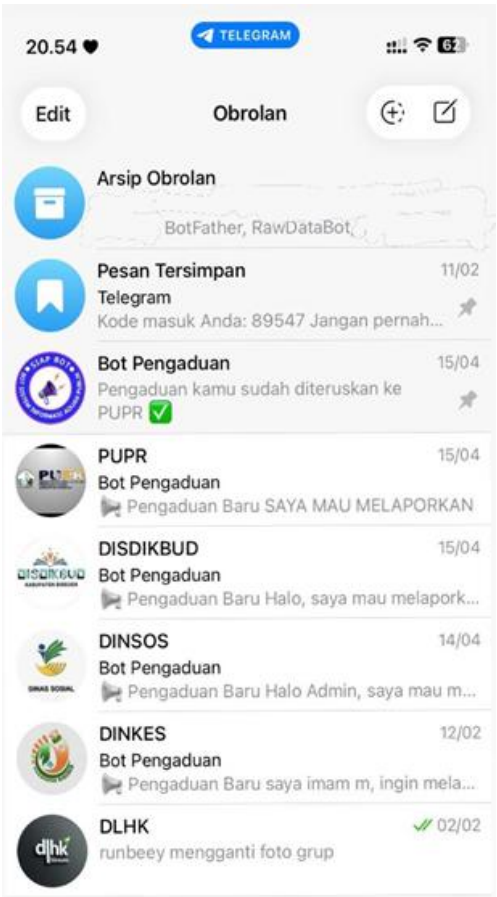
Gambar 2. Confusion Matrix

Berdasarkan confusion matrix pada gambar 2, dapat dilihat bahwa hasil pengujian model SVM yang menunjukkan perbandingan antara label sebenarnya (*true label*) dan label hasil prediksi (*predicted label*). Setiap baris merepresentasikan kelas asli, sedangkan setiap kolom menunjukkan hasil prediksi model. Nilai diagonal (dari kiri atas ke kanan bawah) menunjukkan jumlah data yang diklasifikasikan dengan benar. Terlihat bahwa sebagian besar kelas memiliki nilai diagonal sempurna, yaitu **40 data berhasil diprediksi dengan benar** pada kelas 0 (DINKES), 2 (DISDIKBUD), 3 (DLHK), dan 4 (PUPR). Hal ini menunjukkan bahwa model memiliki kemampuan yang sangat baik dalam mengenali pola pada kelas-kelas tersebut tanpa kesalahan klasifikasi.

Namun, pada kelas 1 (DINSOS) terlihat adanya kesalahan klasifikasi, di mana dari total 40 data, hanya **37 data yang diprediksi dengan benar,** sementara **1 data salah diklasifikasikan ke kelas 2** dan **2 data ke kelas 3.** Kesalahan ini menunjukkan bahwa terdapat kemiripan karakteristik teks antara kelas 1 dengan kelas 2 dan 3, sehingga model mengalami kesulitan dalam membedakan beberapa data. Meskipun demikian, secara keseluruhan performa model tetap sangat baik karena mayoritas data berada pada diagonal utama, yang menandakan tingkat akurasi tinggi dan kesalahan yang relatif kecil.

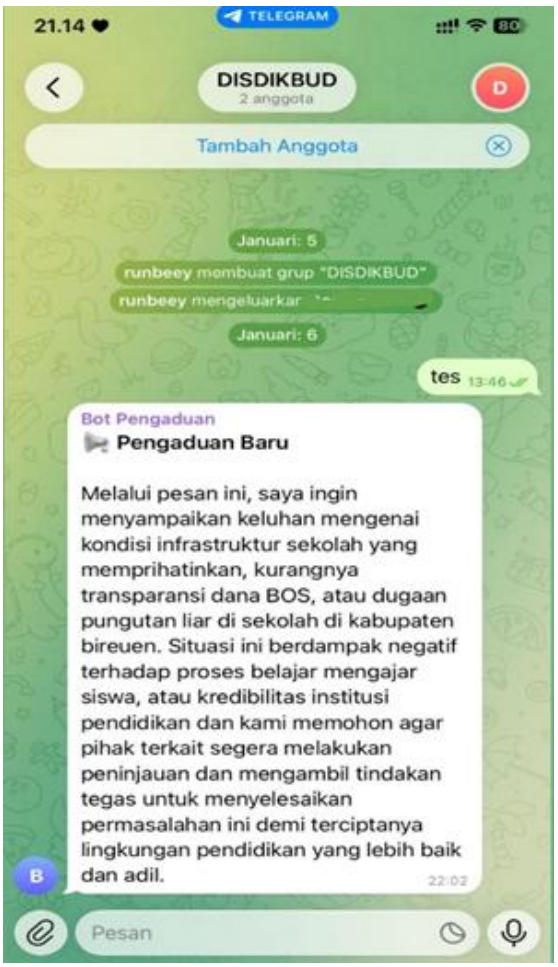
G. Implementasi Sistem

Model klasifikasi yang berhasil dikembangkan kemudian diimplementasikan ke dalam sistem berbasis bot Telegram. Implementasi ini memungkinkan pengguna untuk mengirimkan keluhan secara otomatis dan secara real-time melalui aplikasi Telegram. Program yang dikembangkan ini beroperasi melalui alur kerja yang saling terhubung; prosesnya dimulai ketika seorang pengguna mengirimkanpesan keluhan melalui Telegram, yang kemudian diterima oleh sistem menggunakan API Telegram. Integrasi model pembelajaran mesin dengan bot Telegram dilakukan dengan memanfaatkan model yang telah dilatih sebelumnya dan disimpan sebagai berkas pickle. Model Support Vector Machine (SVM), bersama dengan vektorisasi TF-IDF, dimuat kembali ke dalam sistem sehingga keduanya dapat digunakan secara langsung tanpa perlu dilatih ulang. Ketika sistem beroperasi, setiap pesan masuk dari pengguna terlebih dahulu melalui tahap prapemrosesan sederhana, kemudian dikonversi menjadi vektor numerik melalui TF-IDF, dan selanjutnya diprediksi menggunakan model SVM untuk menentukan kategori keluhan yang sesuai.



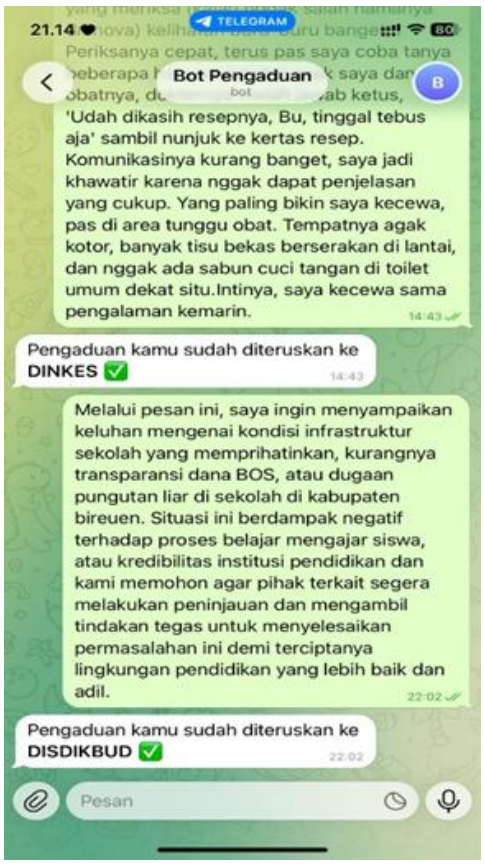
Gambar 3. Tampilan Telegram

Setelah proses klasifikasi selesai, hasil prediksi digunakan sebagai acuan untuk menentukan tujuan distribusi pesan-khususnya, ke grup lembaga pemerintah yang relevan berdasarkan label yang diperoleh. Dalam hal ini, sistem memanfaatkan pemetaan antara label klasifikasi dan ID grup Telegram, sehingga pesan keluhan dapat diteruskan secara otomatis ke instansi yang dituju. Selain itu, bot memberikan respons kepada pengguna dalam bentuk notifikasi yang menyatakan bahwa keluhan telah diproses. Melalui integrasi ini, sistem dapat beroperasi sepenuhnya secara otomatis, mulai dari tahap penerimaan pesan hingga proses pengalihan keluhan, tanpa memerlukan intervensi manual.



Gambar 4. Tampilan Pesan Berhasil Diteruskan Ke Telegram Dinas Terkait

Gambar 4 menunjukkan tampilan pesan pengaduan yang berhasil diteruskan oleh sistem ke grup Telegram dinas terkait, yaitu DISDIKBUD. Pada gambar tersebut terlihat bahwa pesan yang dikirimkan pengguna telah diterima dalam grup tujuan dengan format "Pengaduan Baru", sehingga pihak dinas dapat langsung membaca isi laporan yang masuk. Isi pesan pengaduan berkaitan dengan permasalahan di lingkungan pendidikan, seperti kondisi infrastruktur sekolah, transparansi dana BOS, dugaan pungutan liar, serta permohonan agar pihak terkait melakukan peninjauan dan tindak lanjut. Hal ini menunjukkan bahwa model klasifikasi yang terintegrasi dengan bot Telegram mampu mengenali kategori pengaduan dan meneruskan laporan secara otomatis ke grup instansi yang sesuai. Dengan adanya fitur ini, proses distribusi pengaduan menjadi lebih cepat, terarah, dan mengurangi ketergantungan terhadap penyaluran laporan secara manual.



Gambar 5. Tampilan Pengiriman Pesan Pengaduan oleh Pengguna dan Respon Bot

Gambar 5 menunjukkan tampilan proses pengiriman pesan pengaduan oleh pengguna melalui bot Telegram beserta respons otomatis yang diberikan oleh sistem. Pada gambar tersebut terlihat bahwa pengguna mengirimkan dua pesan pengaduan dengan isi yang berbeda, yaitu keluhan terkait pelayanan kesehatan dan keluhan terkait bidang pendidikan. Setelah pesan dikirimkan, bot secara otomatis memberikan balasan bahwa pengaduan telah diteruskan ke instansi yang sesuai, seperti DINKES untuk pengaduan kesehatan dan DISDIKBUD untuk pengaduan pendidikan. Hal ini menunjukkan bahwa sistem tidak hanya mampu menerima pesan dari pengguna, tetapi juga dapat melakukan klasifikasi isi pengaduan berdasarkan konteks teks, menentukan kategori instansi tujuan, dan memberikan konfirmasi kepada pengguna secara real-time.

KESIMPULAN

Berdasarkan analisis alur kerja pemrosesan data, penerapan metode klasifikasi, dan evaluasi kinerja sistem yang telah dilakukan, dapat disimpulkan bahwa sistem klasifikasi pengaduan publik berbasis Telegram Bot telah berhasil diimplementasikan dengan alur kerja yang terorganisir secara sistematis. Sistem yang dikembangkan mencakup serangkaian tahap, termasuk prapemrosesan data teks, penentuan bobot kata menggunakan metode TF-IDF, klasifikasi dengan algoritma Support Vector Machine (SVM), dan pendistribusian otomatis keluhan ke kelompok instansi pemerintah yang relevan.

Langkah-langkah prapemrosesan yang diterapkan terbukti efektif dalam meningkatkan kualitas data teks, sehingga mendukung pelaksanaan proses pembobotan dan klasifikasi secara optimal. Metode TF-IDF, dikombinasikan dengan pendekatan unigram dan bigram, berhasil merepresentasikan data teks dalam bentuk numerik yang informatif. Selain itu, model SVM menunjukkan kinerja yang sangat memuaskan, ditandai dengan akurasi pengujian sebesar 98,5%, serta nilai presisi, recall, dan F1-score yang tinggi di seluruh kelas yang diuji.

Berdasarkan hasil evaluasi melalui matriks kebingungan, sebagian besar data diklasifikasikan dengan benar dengan tingkat kesalahan yang relatif rendah dan tanpa indikasi overfitting yang signifikan. Selain itu, pengujian sistem juga menunjukkan bahwa bot Telegram mampu menerima, memproses, dan meneruskan keluhan secara otomatis ke penerima yang dituju. Dengan demikian, sistem

yang dikembangkan dalam penelitian ini dianggap telah berhasil mencapai tujuan yang telah ditetapkan, yaitu meningkatkan kecepatan, akurasi, dan efisiensi proses penanganan keluhan publik.

PENDANAAN
Penelitian ini tidak didukung oleh pendanaan apapun.

KONFLIK KEPENTINGAN
Penulis menyatakan bahwa tidak terdapat konflik kepentingan yang dapat mempengaruhi hasil penelitian maupun interpretasi data dalam artikel ini.

KONTRIBUSI PENULIS
Penulis pertama berperan dalam perumusan konsep penelitian, perancangan metode, serta penulisan naskah awal. Penulis kedua berkontribusi dalam pengumpulan dan pengolahan data serta analisis hasil. Penulis ketiga melakukan validasi hasil, revisi naskah, serta penyempurnaan substansi artikel. Seluruh penulis telah membaca dan menyetujui versi akhir naskah.

DAFTAR PUSTAKA
[1] S. D. Utomo, "Penanganan pengaduan Masyarakat mengenai pelayanan publik," *BISNIS & BIROKRASI: Jurnal Ilmu Administrasi dan Organisasi*, vol. 15, no. 3, p. 8, 2011.
[2] Y. Sansena, "Implementasi Sistem Layanan Pengaduan Masyarakat Kecamatan Medan Amplas Berbasis Website," *Jurnal Ilmiah Teknologi Informasi Asia*, vol. 15, no. 2, pp. 91–102, 2021.
[3] H. Stepanus and W. Subadi, "Efektivitas Sistem Pengelolaan Pengaduan Pelayanan Publik Nasional-Layanan Aspirasi Dan Pengaduan Online Rakyat (Sp4n-Lapor) Di Kabupaten Tabalong dalam Penanganan Pengaduan Masyarakat Terkait Pelayanan Publik," *JAPB*, vol. 7, no. 2, pp. 1156–1167, 2024.
[4] M. Haekal and M. B. Z. Tjenreng, "Analisis Pengelolaan Pengaduan Masyarakat melalui Aplikasi JAKI dalam Rangka Pelayanan Publik di DKI Jakarta," *Jurnal PKM Manajemen Bisnis*, vol. 5, no. 1, pp. 228–241, 2025.
[5] D. F. Kuncoro, U. Juniarti, J. Syahputra, R. B. B. Sumantri, and R. Suryani, "Rancang Bangun Sistem Pengaduan Masyarakat Berbasis Web Dengan Metode Waterfall: Array," *Jurnal Sistem Informasi Dan Teknologi Peradaban*, vol. 3, no. 2, pp. 14–19, 2022.
[6] D. Antons, E. Grünwald, P. Cichy, and T. O. Salge, "The application of text mining methods in innovation research: current state, evolution patterns, and development priorities," *R&D Management*, vol. 50, no. 3, pp. 329–351, 2020.
[7] K.-X. Han, W. Chien, C.-C. Chiu, and Y.-T. Cheng, "Application of support vector machine (SVM) in the sentiment analysis of twitter dataset," *Applied Sciences*, vol. 10, no. 3, p. 1125, 2020.
[8] M. Jamaluddin and A. D. Wibawa, "Patient diagnosis classification based on electronic medical record using text mining and support vector machine," in *2021 international seminar on application for Technology of Information and Communication (iSemantic)*, IEEE, 2021, pp. 243–248.
[9] Z. Amanda, I. Muslem, and F. Rizani, "Klasifikasi Kematangan Buah Pepaya Menggunakan Algoritma Support Vector Machine," *Jurnal Ilmu Komputer Aceh*, vol. 3, no. 1, pp. 96–102, Feb. 2026.
[10] B. Laoli, I. Muslem, and F. Rizani, " Klasifikasi Tingkat Kematangan Buah Pisang Menggunakan Algoritma Support Vector Machine ," *Jurnal Ilmu Komputer Aceh*, vol. 3, no. 1, pp. 142–151, Feb. 2026.
[11] M. Furqan, R. A. Putri, I. Agus, and M. Daulay, "IMPLEMENTASI TEXT MINING DALAM IDENTIFIKASI CYBER HATE," vol. 8, no. 4, pp. 477–483, 2024.
[12] I. R. Muslem, "KLIK: Kajian Ilmiah Informatika dan Komputer Image Classification pada Kasus American Sign Language Menggunakan Support Vector Machine," *Media Online)*, vol. 4, no. 2, pp. 1184–1191, 2023, doi: 10.30865/klik.v4i2.1242.
[13] D. Z. Robert, A. B. Setiawan, and D. W. Widodo, "Implementasi Metode Support Vector Machine untuk Klasifikasi Otomatis Pengaduan Publik di Kabupaten Trenggalek," vol. 9, pp. 1617–1626.
[14] D. Teknik, S. Vokasi, and U. G. Mada, "Implementasi Chatbot pada Telegram sebagai Monitoring Assistant dengan Analisis Teks Klasifikasi Menggunakan Metode Support Vector Machine," vol. 5, no. 2, pp. 68–74, 2024.
[15] D. Siregar, F. Ladayya, N. Z. Albaqi, and B. M. Wardana, "Penerapan Metode Support Vector Machines (SVM) dan Metode Naïve Bayes Classifier (NBC) dalam Analisis Sentimen Publik terhadap Konsep Child-free di Media Sosial Twitter," vol. 7, no. 1, pp. 93–104, 2023.
[16] J. Saputra, L. Maryani, D. Wulandari, W. Eka, P. T. Informatika, and P. T. Komputer, "Analisis Performa Naive Bayes dan SVM terhadap Sentimen Teks Media Sosial dengan Word2Vec dan SMOTE," vol. 10, no. 1, pp. 143–155, 2025.

[17] A. Suharman, M. K. Sulaeman, T. Industri, U. Muhammadiyah, and P. Hamka, "Analisis Sentimen Pengguna Aplikasi Livin ' by Mandiri Menggunakan Metode Support Vector Machine (SVM) dengan Ekstraksi Fitur TF-IDF dan Word2Vec User Sentiment Analysis of the Livin ' by Mandiri Application Using the Support Vector Machine (SVM) Method with TF-IDF and Word2Vec Feature Extraction," vol. 5, no. 8, pp. 2201–2212, 2025.

[18] A. Karami, M. Lundy, F. Webb, and Y. K. Dwivedi, "Twitter and research: A systematic literature review through text mining," *IEEE access*, vol. 8, pp. 67698–67717, 2020.

[19] R. Antonius, A. R. Zulkarnain, and H. Irsyad, "Pendekatan TF-IDF , SMOTE , dan SVM dalam Klasifikasi Sentimen Masyarakat terhadap Pemblokiran Judi Online," vol. 2, no. 3, pp. 115–122, 2024, doi: 10.58369/biit.v2i3.65.

[20] A. S. Putri, A. F. Rozi, U. Mercu, B. Yogyakarta, and D. I. Yogyakarta, "SENTIMENT ANALYSIS OF TELEGRAM APPLICATION USER SATISFACTION ON GOOGLE PLAY STORE USING NAÏVE BAYES , LOGISTIC REGRESSION AND SVM ANALISIS SENTIMEN KEPUASAN PENGGUNA APLIKASI TELEGRAM PADA GOOGLE PLAY STORE MENGGUNAKAN METODE NAÏVE BAYES , LOGISTIC REGRESSION DAN SVM," vol. 10, no. 2, pp. 857–867, 2025.

[21] J. Riset, W. Silalahi, and A. Hartanto, "Klasifikasi Sentimen Support Vector Machine Berbasis Optimasi Menyambut Pemilu 2024 Support Vector Machine Sentiment Classification Based on Optimization Welcoming 2024 Election," vol. 7, no. 2, pp. 245–255, 2024, doi: 10.30595/jrst.v7i2.18133.

[22] F. Romadoni, Y. Umaidah, and B. N. Sari, "Text Mining Untuk Analisis Sentimen Pelanggan Terhadap Layanan Uang Elektronik Menggunakan Algoritma Support Vector Machine," vol. 09, pp. 247–253, 2020.

[23] I. Ayu and M. Cahya, "IMPLEMENTASI TEXT MINING UNTUK KLASIFIKASI OPINI ALUMNI PADA PERGURUAN TINGGI," pp. 283–290, 2019.

[24] J. Informatika and S. Informasi, "INFORMASI (Jurnal Informatika dan Sistem Informasi) Volume 12 No.1 / Mei/ 2020," vol. 12, no. 1, pp. 67–80, 2020.