



Department of Digital Business

Journal of Artificial Intelligence and Digital Business (RIGGS)

Homepage: <https://journal.ilmudata.co.id/index.php/RIGGS>

Vol. 5 No. 2 (2026) pp: 11749-11759

P-ISSN: 2963-9298, e-ISSN: 2963-914X

Sistem Informasi Analisis Kebutuhan Fasilitas Publik Berdasarkan Laporan Masyarakat Menggunakan Metode *Text Mining* dan *Support Vector Machine*

Dela Puspita Helmi, Jumadil Nangi, Asa Hari Wibowo

Jurusan Informatika, Fakultas Teknik, Universitas Haluoleo

dhellahelmi@gmail.com, jumadilnangi@uho.ac.id, asa.hari@uho.ac.id

Abstrak

Pengelolaan laporan masyarakat secara manual masih menjadi kendala dalam proses identifikasi kebutuhan fasilitas publik karena memerlukan waktu yang relatif lama, berpotensi menimbulkan ketidakkonsistenan dalam pengelompokan data, serta menyulitkan proses penyusunan prioritas pembangunan. Penelitian ini bertujuan mengembangkan sistem informasi berbasis web yang mampu mengklasifikasikan laporan masyarakat secara otomatis berdasarkan kategori fasilitas publik. Metode yang diterapkan menggunakan Text Mining dengan tahapan case folding, cleaning, dan normalisasi sebagai proses preprocessing, kemudian dilanjutkan dengan pembobotan Term Frequency-Inverse Document Frequency (TF-IDF) serta klasifikasi menggunakan algoritma Support Vector Machine (SVM). Dataset penelitian bersumber dari laporan Musyawarah Perencanaan Pembangunan (Musrenbang) Kabupaten Konawe Tahun 2024 yang terdiri atas 284 data training dan 71 data testing. Evaluasi model dilakukan menggunakan Confusion Matrix dan Classification Report untuk mengukur nilai accuracy, precision, recall, macro F1-score, dan weighted F1-score. Hasil pengujian menunjukkan bahwa model mampu mengklasifikasikan laporan masyarakat ke dalam lima belas kategori fasilitas publik dengan nilai accuracy sebesar 95,77%, precision sebesar 93,54%, recall sebesar 97,04%, macro F1-score sebesar 94,34%, dan weighted F1-score sebesar 96,02%. Hasil tersebut menunjukkan bahwa kombinasi metode TF-IDF dan algoritma Support Vector Machine (SVM) mampu menghasilkan klasifikasi yang baik serta mendukung Bappeda Kabupaten Konawe dalam mengidentifikasi kebutuhan fasilitas publik secara lebih cepat, konsisten, objektif, dan berbasis data sebagai pendukung pengambilan keputusan dalam perencanaan pembangunan daerah.

Kata kunci: Text Mining, Support Vector Machine, TF-IDF, Klasifikasi Teks, Fasilitas Publik, Musrenbang, Sistem Informasi.

1. Latar Belakang

Penyediaan fasilitas publik merupakan salah satu aspek penting dalam mewujudkan pembangunan daerah yang berkelanjutan [1]. Informasi mengenai kebutuhan masyarakat menjadi dasar dalam penyusunan kebijakan dan penentuan prioritas pembangunan. Salah satu sumber informasi tersebut berasal dari laporan masyarakat yang dihimpun melalui kegiatan Musyawarah Perencanaan Pembangunan (Musrenbang) [2]. Namun, laporan masyarakat umumnya berbentuk teks tidak terstruktur dengan variasi penggunaan istilah, bahasa, dan cara penyampaian yang berbeda sehingga proses identifikasi kebutuhan fasilitas publik secara manual menjadi kurang efektif dan memerlukan waktu yang relatif lama. Selain itu, perkembangan transformasi digital mendorong pemerintah untuk memanfaatkan teknologi informasi dalam mendukung proses perencanaan pembangunan berbasis data [3].

BAPPEDA Kabupaten Konawe menghadapi permasalahan serupa dalam mengelola laporan hasil Musrenbang Tahun 2024 [4]. Banyaknya usulan masyarakat yang mencakup berbagai kebutuhan fasilitas publik, seperti infrastruktur jalan, jembatan, drainase, fasilitas pendidikan, fasilitas kesehatan, irigasi pertanian, dan penyediaan air bersih, masih dikelompokkan secara manual. Perbedaan penulisan, penggunaan sinonim, maupun istilah yang memiliki makna serupa menyebabkan proses klasifikasi membutuhkan waktu yang lama dan berpotensi menghasilkan ketidakkonsistenan dalam pengelompokan laporan. Kondisi tersebut menghambat proses identifikasi kebutuhan fasilitas publik serta penyusunan prioritas pembangunan daerah secara tepat dan efisien [5].

Perkembangan teknologi *text mining* memberikan solusi dalam mengolah data teks tidak terstruktur menjadi informasi yang dapat dimanfaatkan untuk mendukung pengambilan keputusan [6]. Melalui tahapan *case folding*, *cleaning*, normalisasi dan pembobotan TF-IDF, informasi penting pada laporan masyarakat dapat direpresentasikan dalam bentuk numerik sehingga dapat diproses menggunakan algoritma klasifikasi. Salah satu algoritma yang banyak digunakan adalah *Support Vector Machine* (SVM) karena memiliki kemampuan yang baik dalam mengklasifikasikan data teks berdimensi tinggi serta menghasilkan performa klasifikasi yang tinggi pada berbagai penelitian[7].

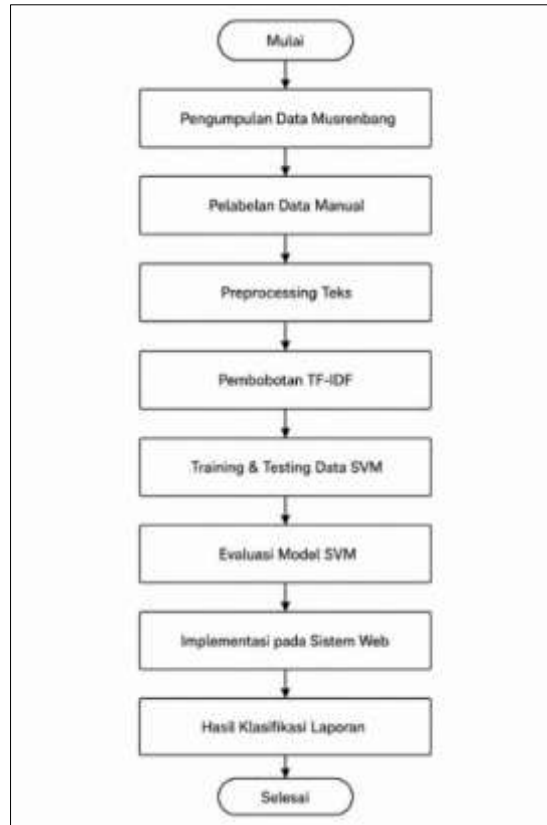
Beberapa penelitian terdahulu telah menerapkan text mining dan algoritma SVM untuk mengklasifikasikan pengaduan maupun opini masyarakat. Penelitian Limansyah dkk. memanfaatkan *text mining* untuk menganalisis kebutuhan pengunjung pusat perbelanjaan [8], sedangkan Rasywir dkk. mengklasifikasikan sentimen keluhan masyarakat pada media sosial [9]. Muqorobin dkk. juga menerapkan klasifikasi terhadap pengaduan layanan publik menggunakan data Twitter[10]. Hasil penelitian tersebut menunjukkan bahwa metode *text mining* mampu meningkatkan efektivitas pengolahan data teks. Akan tetapi, sebagian besar penelitian masih berfokus pada performa model klasifikasi dan belum mengintegrasikan hasil klasifikasi ke dalam sistem informasi yang mendukung proses perencanaan pembangunan daerah.

Berdasarkan permasalahan tersebut, penelitian ini mengembangkan sistem informasi berbasis web untuk mengklasifikasikan laporan masyarakat ke dalam lima belas kategori kebutuhan fasilitas publik menggunakan tahapan *text mining*, pembobotan TF-IDF, dan algoritma *Support Vector Machine*. Hasil klasifikasi disajikan dalam bentuk tabel dan visualisasi sehingga memudahkan BAPPEDA Kabupaten Konawe dalam mengidentifikasi distribusi kebutuhan fasilitas publik secara lebih cepat, objektif, dan berbasis data. Kebaruan penelitian ini terletak pada integrasi proses klasifikasi otomatis dengan sistem informasi yang dapat mendukung penyusunan prioritas pembangunan berdasarkan laporan masyarakat.

2. Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan menerapkan teknik text mining dan algoritma Support Vector Machine (SVM) untuk mengklasifikasikan laporan masyarakat berdasarkan kebutuhan fasilitas publik. Data yang digunakan merupakan laporan hasil Musyawarah Perencanaan Pembangunan (Musrenbang) Kabupaten Konawe Tahun 2024 yang diperoleh dari Badan Perencanaan Pembangunan Daerah (BAPPEDA) Kabupaten Konawe. Seluruh data telah melalui proses pelabelan ke dalam lima belas kategori fasilitas publik yang meliputi Infrastruktur Jalan, Jembatan, Drainase, Air Bersih dan SPAM, Irigasi Pertanian, Fasilitas Pendidikan, Fasilitas Kesehatan, Rumah Ibadah, Penerangan Jalan Umum, Pengelolaan Sampah, Transportasi Umum, Sarana Olahraga, Ruang Terbuka Publik, Sanitasi Lingkungan, dan Tidak Terklasifikasi.

Tahapan penelitian dilakukan secara berurutan, dimulai dari pengumpulan data, preprocessing teks, pembobotan menggunakan Term Frequency–Inverse Document Frequency (TF-IDF), pelatihan model menggunakan algoritma SVM, proses klasifikasi laporan, hingga evaluasi kinerja model menggunakan confusion matrix, accuracy, precision, recall, dan F1-score. Alur penelitian yang diterapkan pada penelitian ini ditunjukkan pada Gambar 1.



Gambar 1. Tahapan Penelitian

2.1. Data Penelitian

Data yang digunakan dalam penelitian ini merupakan laporan masyarakat hasil Musyawarah Perencanaan Pembangunan (Musrenbang) Kabupaten Konawe Tahun 2024 yang diperoleh dari BAPPEDA Kabupaten Konawe. Data berupa teks tidak terstruktur yang memuat usulan pembangunan fasilitas publik dari berbagai wilayah di Kabupaten Konawe.

Sebelum digunakan dalam proses klasifikasi, setiap laporan diberi label secara manual ke dalam lima belas kategori fasilitas publik. Selanjutnya, data dibagi menjadi data *training* dan data *testing* sebagai dasar pelatihan serta evaluasi model *Support Vector Machine* (SVM). Distribusi data penelitian berdasarkan kategori ditunjukkan pada Tabel 1.

Tabel 1. Data Penelitian

No	Kategori Fasilitas Publik	Jumlah Data
1	Infrastruktur Jalan	76
2	Jembatan	25
3	Drainase	62
4	Air Bersih dan SPAM	27
5	Irigasi Pertanian	40
6	Fasilitas Pendidikan	21
7	Fasilitas Kesehatan	20
8	Rumah Ibadah	41
9	Penerangan Jalan Umum	8
10	Pengelolaan Sampah	8
11	Transportasi Umum	12
12	Sarana Olahraga	11
13	Ruang Terbuka Publik	8
14	Sanitasi Lingkungan	28
15	Tidak Terklasifikasi	12
Total		536

Berdasarkan Tabel 1, data penelitian terdiri dari 536 laporan masyarakat yang telah dikelompokkan ke dalam 15 kategori fasilitas publik. Kategori dengan jumlah data terbanyak adalah Infrastruktur Jalan sebanyak 76 laporan, sedangkan kategori Penerangan Jalan Umum, Pengelolaan Sampah, dan Ruang Terbuka Publik masing-masing memiliki 8 laporan. Distribusi data tersebut digunakan sebagai dasar dalam proses pelatihan (*training*) dan pengujian (*testing*) model klasifikasi.

2.2. Tahapan Penelitian

Tahapan penelitian dimulai dengan pengumpulan data laporan masyarakat hasil Musrenbang Kabupaten Konawe Tahun 2024 yang telah diberi label sesuai kategori fasilitas publik. Selanjutnya, data diproses melalui tahapan preprocessing teks yang meliputi *case folding*, *cleaning*, dan normalisasi. Hasil preprocessing kemudian dibobot menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF) sebelum dilakukan proses pelatihan (*training*) dan pengujian (*testing*) menggunakan algoritma *Support Vector Machine* (SVM). Tahap akhir penelitian adalah mengevaluasi kinerja model menggunakan *confusion matrix*, *accuracy*, *precision*, *recall*, dan *F1-score*, kemudian mengimplementasikan model ke dalam sistem informasi berbasis web untuk mengklasifikasikan laporan masyarakat secara otomatis.

2.3. Preprocessing Teks

Preprocessing teks merupakan tahap awal dalam proses *text mining* yang bertujuan untuk meningkatkan kualitas data dengan mengubah teks tidak terstruktur menjadi bentuk yang lebih konsisten sehingga dapat diproses pada tahap klasifikasi. Pada penelitian ini, tahapan *preprocessing* terdiri atas *case folding*, *cleaning*, dan normalisasi.

1. Case Folding
Case folding merupakan proses mengubah seluruh karakter pada teks menjadi huruf kecil (*lowercase*) untuk menyeragamkan penulisan kata. Tahapan ini bertujuan menghilangkan perbedaan antara huruf kapital dan huruf kecil sehingga kata yang memiliki makna sama dapat dikenali sebagai satu istilah yang konsisten dalam proses pengolahan data [11].
2. Cleaning
Cleaning merupakan proses menghapus karakter yang tidak diperlukan, seperti tanda baca, angka, simbol, dan karakter khusus lainnya sehingga hanya menyisakan kata-kata yang memiliki makna dalam proses klasifikasi [12].
3. Normalisasi
Normalisasi merupakan proses mengubah kata tidak baku, singkatan, atau istilah yang memiliki penulisan berbeda menjadi bentuk kata baku [13].

2.4. Term Frequency-Inverse Document Frequency (TF-IDF)

Metode TF-IDF digunakan untuk menentukan tingkat kepentingan suatu kata dalam sebuah dokumen berdasarkan frekuensi kemunculannya pada dokumen tersebut serta jumlah dokumen yang memuat kata yang sama [14]. Semakin sering suatu kata muncul pada satu dokumen dan semakin sedikit muncul pada dokumen lain, maka bobot kata tersebut akan semakin besar.

Pembobotan TF-IDF dihitung menggunakan Persamaan (1) dan Persamaan (2).

$$TF - IDF(t, d) = TF(t, d) \times IDF(t) \quad (1)$$

$$IDF(t) = \log\left(\frac{N}{df}\right) \quad (2)$$

dengan :

- a) TF = frekuensi kemunculan kata pada suatu dokumen,
- b) IDF = bobot kebalikan frekuensi dokumen,
- c) N = jumlah seluruh dokumen,
- d) df = jumlah dokumen yang mengandung kata.

2.5. Support Vector Machine (SVM)

Algoritma *Support Vector Machine* (SVM) merupakan metode klasifikasi *supervised learning* yang membentuk *hyperplane* optimal untuk memisahkan data berdasarkan kelasnya. Pada penelitian ini, SVM digunakan untuk mengklasifikasikan laporan masyarakat yang telah direpresentasikan dalam bentuk vektor TF-IDF ke dalam lima belas kategori fasilitas public [15].

Fungsi keputusan pada algoritma SVM dinyatakan pada Persamaan (5).

$$f(x) = w \cdot x + b \quad (5)$$

dengan:

- a) $f(x)$ = hasil fungsi keputusan.
- b) w = vektor bobot (*weight vector*).
- c) x = vektor fitur hasil pembobotan TF-IDF.
- d) b = nilai *bias*.

2.6. Evaluasi Model

Kinerja model dievaluasi menggunakan *confusion matrix*, yaitu tabel yang membandingkan hasil prediksi model dengan label sebenarnya pada data *testing*. Berdasarkan *confusion matrix*, dihitung beberapa metrik evaluasi, yaitu *accuracy*, *precision*, *recall*, dan *F1-score* [15].

Perhitungan *accuracy* ditunjukkan pada Persamaan (6).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

Perhitungan *precision* ditunjukkan pada Persamaan (7).

$$Precision = \frac{TP}{TP+FP} \quad (7)$$

Perhitungan *recall* ditunjukkan pada Persamaan (8).

$$Recall = \frac{TP}{TP+FN} \quad (8)$$

Perhitungan *F1-score* ditunjukkan pada Persamaan (9).

$$F1-Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (9)$$

dengan:

- a) TP (*True Positive*) = jumlah data yang berhasil diklasifikasikan dengan benar.
- b) TN (*True Negative*) = jumlah data yang berhasil dikenali sebagai bukan kelas tersebut.
- c) FP (*False Positive*) = jumlah data yang diprediksi sebagai suatu kelas, tetapi sebenarnya berasal dari kelas lain.
- d) FN (*False Negative*) = jumlah data yang seharusnya berada pada suatu kelas, tetapi diprediksi sebagai kelas lain.

Evaluasi menggunakan metrik tersebut bertujuan untuk memberikan gambaran menyeluruh mengenai kemampuan model dalam mengklasifikasikan laporan masyarakat berdasarkan kategori fasilitas publik. Nilai evaluasi yang diperoleh selanjutnya digunakan untuk menganalisis performa model pada bagian hasil dan pembahasan.

3. Hasil dan Diskusi

Bagian ini menyajikan hasil implementasi sistem informasi analisis kebutuhan fasilitas publik berdasarkan laporan masyarakat menggunakan metode Text Mining dan algoritma Support Vector Machine (SVM). Pembahasan meliputi hasil implementasi sistem berbasis web, proses klasifikasi laporan masyarakat, serta evaluasi kinerja model menggunakan *confusion matrix*, *accuracy*, *precision*, *recall*, dan *F1-score*. Hasil yang diperoleh digunakan untuk mengetahui kemampuan sistem dalam mengklasifikasikan laporan masyarakat secara otomatis sesuai dengan kategori fasilitas publik yang telah ditentukan.

3.1. Hasil *Preprocessing* Teks

Tahap *preprocessing* dilakukan untuk mempersiapkan data laporan masyarakat sebelum proses pembobotan TF-IDF dan klasifikasi menggunakan algoritma Support Vector Machine (SVM). Pada penelitian ini, *preprocessing* terdiri atas tiga tahapan, yaitu *case folding*, *cleaning*, dan *normalisasi*. Hasil dari setiap tahapan ditunjukkan sebagai berikut.

1. Case Folding

Proses *case folding* dilakukan dengan mengubah seluruh huruf pada laporan masyarakat menjadi huruf kecil (*lowercase*) untuk menyeragamkan penulisan kata. Tahapan ini bertujuan agar tidak terdapat perbedaan antara huruf kapital dan huruf kecil dalam proses pengolahan data. Contoh hasil proses *case folding* ditunjukkan pada Tabel 2.

Tabel 2. Tahapan *Case Folding*

Sebelum	Sesudah
Jalan Desa Asinua Tua Masih Rusak Berat dan Belum Pernah Diperbaiki	jalan desa asinua tua masih rusak berat dan belum pernah diperbaiki

2. Cleaning

Proses *cleaning* dilakukan dengan menghapus karakter yang tidak diperlukan, seperti tanda baca, angka, dan simbol, sehingga menghasilkan teks yang lebih bersih tanpa mengubah makna laporan. Contoh hasil proses *cleaning* ditunjukkan pada Tabel 3.

Tabel 3. Tahapan Proses *Case Folding*

Sebelum	Sesudah
Jalan desa asinua tua masih rusak!!! mohon segera diperbaiki, 2024.	jalan desa asinua tua masih rusak mohon segera diperbaiki

3. Normalisasi

Tahapan ini bertujuan untuk menyeragamkan representasi kata sehingga kata yang memiliki makna sama dapat dikenali sebagai satu istilah yang seragam dan meningkatkan akurasi proses klasifikasi.

Tabel 4. Tahapan Proses Normalisasi

Sebelum	Sesudah
jalan desa asinua tua msh rusak, blm pernah diperbaiki krn aksesnya susah	jalan desa asinua tua masih rusak, belum pernah diperbaiki karena aksesnya susah

3.2. Pembobotan TF-IDF

Setelah proses *preprocessing* selesai, setiap kata pada laporan masyarakat diberikan bobot menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF). Sebagai contoh, digunakan tiga dokumen hasil *preprocessing* seperti ditunjukkan pada Tabel 5.

Tabel 5. Contoh Dokumen Hasil *Preprocessing*

Dokumen	Isi Dokumen
D1	jalan rusak desa asinua
D2	Jalan Jembatan rusak
D3	irigasi pertanian rusak

Sebagai contoh, dilakukan perhitungan bobot untuk kata jalan. Jumlah dokumen (N) adalah 3, kata jalan muncul pada 2 dokumen (df = 2), dan frekuensi kemunculan kata jalan pada dokumen D1 adalah 1.

Nilai IDF dihitung menggunakan Persamaan (10).

$$IDF(jalan) = \log\left(\frac{3}{2}\right) = 0,176 \quad (10)$$

Selanjutnya, bobot TF-IDF dihitung menggunakan Persamaan (11).

$$W(jalan, D1) = TF(jalan, D1) \times IDF(jalan) \quad (11)$$

Karena nilai TF(jalan,D1) = 1, maka bobot TF-IDF diperoleh sebagai berikut.

$$W(jalan, D1) = 1 \times 0,176 \quad (12)$$

Hasil perhitungan menunjukkan bahwa kata jalan memiliki bobot sebesar 0,176 pada dokumen D1. Proses yang sama diterapkan pada seluruh kata dalam setiap laporan sehingga menghasilkan representasi vektor numerik yang selanjutnya digunakan sebagai masukan pada proses klasifikasi menggunakan algoritma *Support Vector Machine* (SVM).

3.3. Klasifikasi Menggunakan Support Vector Machine

Hasil pembobotan TF-IDF berupa vektor numerik digunakan sebagai masukan bagi algoritma Support Vector Machine (SVM) untuk mengklasifikasikan laporan masyarakat ke dalam lima belas kategori fasilitas publik. Pada tahap ini, model SVM mempelajari pola dari data training kemudian menentukan kategori yang paling sesuai untuk setiap laporan pada data testing.

Selain menghasilkan kategori prediksi, sistem juga menampilkan nilai *confidence score* yang menunjukkan tingkat keyakinan model terhadap hasil klasifikasi. Pada penelitian ini diterapkan nilai ambang (*threshold*) sebesar 0,50. Apabila nilai *confidence* $\geq$ 0,50, maka hasil prediksi diterima sebagai kategori fasilitas publik. Sebaliknya, apabila nilai *confidence* $<$ 0,50, maka laporan dikategorikan sebagai Tidak Terklasifikasi sehingga dapat ditinjau kembali oleh administrator.

Tabel 6. Contoh Hasil Klasifikasi SVM

Laporan	Kategori Prediksi	Confidence	Hasil
Jalan desa Asinua rusak dan berlubang	Infrastruktur Jalan	0,64	Terklasifikasi
Jembatan penghubung mulai rusak	Jembatan	0,70	Terklasifikasi
Mohon dibangun tempat olahraga	Sarana Olahraga	0,63	Terklasifikasi
Jalan menuju kebun perlu perhatian	Infrastruktur Jalan	0,47	Tidak Terklasifikasi

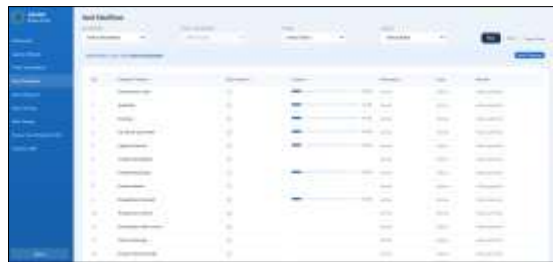
Berdasarkan Tabel 6, sistem hanya menerima hasil klasifikasi dengan nilai *confidence* minimal 0,50. Penerapan *threshold* tersebut bertujuan untuk meningkatkan keandalan hasil klasifikasi sehingga laporan dengan tingkat keyakinan rendah tidak langsung dimasukkan ke dalam kategori tertentu, tetapi dialihkan ke kategori Tidak Terklasifikasi.

3.4. Hasil Implementasi Sistem

Sistem Informasi Analisis Kebutuhan Fasilitas Publik berhasil dikembangkan berbasis web untuk mendukung proses pengelolaan dan klasifikasi laporan masyarakat menggunakan metode *Text Mining* dan algoritma *Support Vector Machine* (SVM). Sistem terdiri atas dua jenis pengguna, yaitu masyarakat dan administrator (admin). Masyarakat dapat mengirimkan laporan kebutuhan fasilitas publik melalui sistem, sedangkan admin bertugas mengelola data, menjalankan proses klasifikasi, serta memantau hasil analisis melalui dashboard.

1. Halaman Hasil Klasifikasi

Halaman hasil klasifikasi menampilkan laporan masyarakat yang telah diproses beserta kategori fasilitas publik dan nilai *confidence* yang dihasilkan oleh model SVM. Sistem juga menyediakan fitur penyaringan data berdasarkan wilayah dan periode pelaporan sehingga memudahkan proses analisis.



Gambar 1. Hasil Klasifikasi

a) Halaman Evualuasi Model SVM

Halaman evaluasi model menyajikan hasil pengujian model klasifikasi dalam bentuk *classification report*, *confusion matrix*, serta nilai *accuracy*. Informasi tersebut digunakan untuk mengevaluasi performa model SVM dalam mengklasifikasikan laporan masyarakat dan menjadi dasar dalam analisis hasil penelitian.



Gambar 2. Halaman Hasil Evaluasi Model SVM

3.2. Hasil Evaluasi Model

Evaluasi model dilakukan untuk mengukur kemampuan algoritma *Support Vector Machine* (SVM) dalam mengklasifikasikan laporan masyarakat berdasarkan kategori fasilitas publik. Pengujian dilakukan menggunakan 71 data testing yang sebelumnya tidak digunakan pada proses pelatihan model. Evaluasi dilakukan menggunakan *confusion matrix* serta metrik *accuracy*, *precision*, *recall*, dan *F1-score*.

a) *Confusion Matrix*

Confusion matrix digunakan untuk mengetahui hubungan antara hasil prediksi model dengan label sebenarnya pada setiap kategori. Melalui matriks ini dapat diketahui jumlah data yang berhasil diklasifikasikan dengan benar maupun data yang mengalami kesalahan klasifikasi.

Aktual / Prediksi	A	D	F	F	I	I	J	P	P	R	R	S	S	T
	B	R	K	P	J	R	B	J	S	T	I	T	O	R
	S	N	S	D	L	G	T	U	P	P	B	L	R	U
ABS	5	0	0	0	0	0	0	0	0	0	0	0	0	0
DRN	0	12	0	0	0	0	0	0	0	0	0	0	0	0
FKS	0	0	4	0	0	0	0	0	0	0	0	0	0	0
FPD	0	0	0	4	0	0	0	0	0	0	0	0	0	0
IIL	0	0	0	0	13	0	0	1	0	0	0	0	0	0
IRG	0	0	0	0	1	6	0	0	0	0	0	0	0	0
JBT	0	0	0	0	0	0	4	0	0	0	0	0	0	0
PJU	0	0	0	0	0	0	0	1	0	0	0	0	0	0
PSP	0	0	0	0	0	0	0	0	1	0	0	0	0	0
RTP	0	0	0	0	0	0	0	0	0	2	0	0	0	0
RIB	0	0	0	0	0	0	0	0	0	0	8	0	0	0
STL	0	0	0	0	0	0	0	0	0	0	0	4	0	1
SOR	0	0	0	0	0	0	0	0	0	0	0	0	2	0
TRU	0	0	0	0	0	0	0	0	0	0	0	0	0	2

Gambar 3. *Confusion Matrix*

Berdasarkan Gambar 3, sebagian besar data uji berhasil diklasifikasikan pada kategori yang sesuai. Hal ini ditunjukkan oleh dominasi nilai pada diagonal utama *confusion matrix*, yang menunjukkan bahwa prediksi model sesuai dengan label sebenarnya. Dari total 71 data *testing*, sebanyak 68 data berhasil diklasifikasikan dengan benar, sedangkan 3 data mengalami kesalahan klasifikasi. Kesalahan tersebut terjadi pada kategori Infrastruktur Jalan, Irigasi Pertanian, dan Sanitasi Lingkungan, yang diprediksi ke kategori lain akibat adanya kemiripan kosakata dan konteks antar laporan. Secara keseluruhan, hasil *confusion matrix* menunjukkan bahwa model SVM memiliki kemampuan klasifikasi yang baik dalam mengelompokkan laporan masyarakat berdasarkan kategori fasilitas publik.

b) *Classification Report*

Untuk memperoleh informasi yang lebih rinci mengenai kinerja model pada setiap kategori fasilitas publik, dilakukan evaluasi menggunakan *classification report*. Evaluasi ini menyajikan nilai *precision*, *recall*, *F1-score*, dan *support* untuk masing-masing kategori sehingga kemampuan model dalam mengklasifikasikan setiap kelas dapat dianalisis secara lebih komprehensif. Hasil *classification report* ditunjukkan pada Tabel 7.

Tabel 7. *Classification Report*

Kategori	Precision	Recall	F1	Support
Air Bersih dan SPAM	100.0%	100.0%	100.0%	5
Drainase	100.0%	100.0%	100.0%	12
Fasilitas Kesehatan	100.0%	100.0%	100.0%	4
Fasilitas Pendidikan	100.0%	100.0%	100.0%	4
Infrastruktur Jalan	92.9%	92.9%	92.9%	14
Irigasi Pertanian	100.0%	85.7%	92.3%	7
Jembatan	100.0%	100.0%	100.0%	4
Penerangan Jalan Umum	50.0%	100.0%	66.7%	1
Pengelolaan Sampah	100.0%	100.0%	100.0%	1
Ruang Terbuka Publik	100.0%	100.0%	100.0%	2
Rumah Ibadah	100.0%	100.0%	100.0%	8
Sanitasi Lingkungan	100.0%	80.0%	88.9%	5

Sarana Olahraga	100.0%	100.0%	100.0%	2
Transportasi Umum	66.7%	100.0%	80.0%	2

Berdasarkan Tabel 7, sebagian besar kategori memperoleh nilai *precision*, *recall*, dan *F1-score* sebesar 100%, yang menunjukkan bahwa model mampu mengklasifikasikan laporan dengan sangat baik. Kategori Infrastruktur Jalan memperoleh nilai *F1-score* sebesar 92,9%, sedangkan Irigasi Pertanian memperoleh 92,3%. Nilai *F1-score* terendah terdapat pada kategori Penerangan Jalan Umum sebesar 66,7% dan Transportasi Umum sebesar 80,0%. Hal ini disebabkan oleh jumlah data pada kategori tersebut yang relatif sedikit sehingga variasi pola laporan yang dipelajari model menjadi terbatas. Secara keseluruhan, hasil *classification report* menunjukkan bahwa model SVM memiliki performa yang baik dalam mengklasifikasikan laporan masyarakat ke dalam kategori fasilitas publik.

c) Hasil Evaluasi Model *Support Vector Machine* (SVM)

Selain evaluasi pada setiap kategori menggunakan *classification report*, kinerja model secara keseluruhan juga diukur berdasarkan nilai *accuracy*, *precision*, *recall*, *macro F1-score*, dan *weighted F1-score*. Metrik tersebut digunakan untuk mengetahui tingkat ketepatan dan kemampuan model dalam mengklasifikasikan laporan masyarakat. Hasil evaluasi model *Support Vector Machine* (SVM) ditunjukkan pada Tabel 8.

Tabel 8. Hasil Evaluasi Model SVM

Metrik	Nilai
Accuracy	95,77%
Precision	93,54%
Recall	97,04%
Macro F1-Score	94,34%
Weighted F1-Score	96,02%

Berdasarkan Tabel 8, model *Support Vector Machine* (SVM) memperoleh nilai *accuracy* sebesar 95,77%, yang menunjukkan bahwa model mampu mengklasifikasikan sebagian besar data *testing* ke dalam kategori fasilitas publik dengan benar. Nilai tersebut menunjukkan bahwa kombinasi tahapan *preprocessing*, pembobotan TF-IDF, dan algoritma SVM mampu mengenali pola pada laporan masyarakat secara efektif.

Nilai *precision* sebesar 93,54% menunjukkan bahwa hasil prediksi yang diberikan model memiliki tingkat ketepatan yang tinggi, sedangkan nilai *recall* sebesar 97,04% menunjukkan bahwa model mampu mengenali sebagian besar laporan sesuai dengan kategori sebenarnya. Selain itu, model memperoleh *macro F1-score* sebesar 94,34% dan *weighted F1-score* sebesar 96,02%, yang menunjukkan keseimbangan yang baik antara *precision* dan *recall* pada seluruh kategori fasilitas publik.

Secara keseluruhan, hasil evaluasi menunjukkan bahwa model SVM memiliki performa yang sangat baik dalam mengklasifikasikan laporan masyarakat. Tingginya nilai pada setiap metrik evaluasi menunjukkan bahwa sistem mampu memberikan hasil klasifikasi yang akurat sehingga dapat mendukung BAPPEDA Kabupaten Konawe dalam mengidentifikasi kebutuhan fasilitas publik secara lebih cepat, konsisten, dan berbasis data.

4. Kesimpulan

Penelitian ini berhasil mengembangkan Sistem Informasi Analisis Kebutuhan Fasilitas Publik berbasis web yang mampu mengklasifikasikan laporan masyarakat secara otomatis menggunakan tahapan Text Mining, pembobotan TF-IDF, dan algoritma *Support Vector Machine* (SVM). Sistem yang dikembangkan dapat membantu BAPPEDA Kabupaten Konawe dalam mengelompokkan laporan masyarakat ke dalam lima belas kategori fasilitas publik sehingga proses pengolahan data menjadi lebih cepat, terstruktur, dan mendukung penyusunan prioritas pembangunan berbasis data. Hasil pengujian menunjukkan bahwa model SVM memiliki performa yang sangat baik dengan nilai *accuracy* sebesar 95,77%, *precision* sebesar 93,54%, *recall* sebesar 97,04%, *macro F1-score* sebesar 94,34%, dan *weighted F1-score* sebesar 96,02%. Hasil tersebut menunjukkan bahwa kombinasi tahapan *preprocessing*, pembobotan TF-IDF, dan algoritma SVM mampu mengklasifikasikan laporan masyarakat dengan tingkat akurasi yang tinggi. Pengembangan penelitian selanjutnya dapat dilakukan dengan menambahkan jumlah data pelatihan yang lebih beragam, menerapkan metode klasifikasi lain sebagai pembanding, serta mengintegrasikan sistem dengan sumber data laporan masyarakat secara *real-time* sehingga dapat meningkatkan

akurasi model dan mendukung proses pengambilan keputusan yang lebih efektif dalam perencanaan pembangunan daerah.

Referensi

1. I. Bastian, *Sistem Pengendalian Manajemen Sektor Publik (Mempertahankan Kepentingan Masyarakat)*. Jakarta: Salemba Empat.
2. Irham, "Partisipasi Pemangku Kepentingan Dalam Musyawarah Perencanaan Pembangunan Di Kecamatan Tempe Kabupaten Wajo," *Jurnal Administrasi Publik*, Vol. 15, No. 2, Hlm. 172–181, Des 2019, Doi: 10.52316/Jap.V15i2.34.
3. G. Ginting, A. Kuswandi, Dan A. Budiati, "Partisipasi Masyarakat Dalam Perencanaan Pembangunan Desa Kandui: Faktor Pengaruh Dan Tantangan," *Neorespublica : Jurnal Ilmu Pemerintahan*, Vol. 6, No. 1, Hlm. 112–124, Des 2024, Doi: 10.52423/Neores.V6i1.1084.
4. Badan Pusat Statistik Kabupaten Konawe, *Statistik Potensi Desa Kabupaten Konawe*. 2024.
5. Badan Pusat Statistik Kabupaten Konawe, "Kabupaten Konawe Dalam Angka," 2025.
6. L. Cahyani Dan M. Arif, "Text Mining Untuk Pengelompokan Skripsi Di Prodi Pendidikan Informatika Universitas Trunojoyo Madura," *Jurnal Ilmiah Edutic : Pendidikan Dan Informatika*, Vol. 8, No. 2, Hlm. 97–108, Jun 2022, Doi: 10.21107/Edutic.V8i2.13020.
7. D. Z. Robert, A. B. Setiawan, Dan D. W. Widodo, "Implementasi Metode Support Vector Machine Untuk Klasifikasi Otomatis Pengaduan Publik Di Kabupaten Trenggalek," Dalam *Prosiding Semnas Inotek (Seminar Nasional Inovasi Teknologi)*, 2025, Hlm. 1617–1624. [Daring]. Tersedia Pada: <https://Proceeding.Unpkediri.Ac.Id/Index.Php/Inotek/>
8. F. Limansyah, Mokh. Suef, Dan V. Ratnasari, "Visitors Needs Analysis In Mall Xyz With Text Mining Analysis," *Iptek Journal Of Proceedings Series*, Vol. 0, No. 1, Hlm. 152, 2021, Doi: 10.12962/J23546026.Y2020i1.11321.
9. E. Rasywir, Y. Pratama, Irawan, Dan M. Istoningtyas, "Public Complaint Tweet Data Feature Analysis For Sentiment Classification," *Bulletin Of Electrical Engineering And Informatics*, Vol. 13, No. 6, Hlm. 4366–4377, 2024, Doi: 10.11591/Eei.V13i6.7172.
10. M. Muqorobin, S. Rokhmah, I. Muslihah, Dan N. A. Rozaq Rais, "Classification Of Community Complaints Against Public Services On Twitter," *International Journal Of Computer And Information System (Ijcis)*, Vol. 1, No. 1, Hlm. 7–10, 2020, Doi: 10.29040/Ijcis.V1i1.6.
11. I. Syarif, R. Asmara, Dan N. U. Rusmayani, "Klasifikasi Keluhan Masyarakat Pada Sosial Media Twitter Terhadap Pelayanan Toko Online Di Indonesia Menggunakan Metode Cosine Tf-Idf," *Bina Insani Ict Journal*, Vol. 7, No. 1, Hlm. 33–42, 2020, [Daring]. Tersedia Pada: <https://Proceeding.Unpkediri.Ac.Id/Index.Php/Inotek/>
12. Y.-C. Chen, C.-H. Chang, Dan Y.-H. Lin, "Mining Public Service Quality Feedback From Social Media: A Computational Analytics Method," *Gov. Inf. Q.*, Vol. 38, No. 2, 2021, Doi: 10.1016/J.Giq.2021.101571.
13. E. D. Madyatmadja, C. P. M. Sianipar, C. Wijaya, Dan D. J. M. Sembiring, "Classifying Crowdsourced Citizen Complaints Through Data Mining: Accuracy Testing Of K-Nearest Neighbors, Random Forest, Support Vector Machine, And," Vol. 10, No. 4, 2023.
14. V. Meida Hersianty, E. Larasati Amalia, D. Puspitasari, Dan D. Wahyu Wibowo, "Penerapan Algoritma Tf-Idf Dan Cosine Similarity Dalam Sistem Rekomendasi Lowongan Pekerjaan," 2025.
15. F. S. Pamungkas Dan I. Kharisudin, "Analisis Sentimen Dengan Svm, Naïve Bayes Dan Knn Untuk Studi Tanggapan Masyarakat Indonesia Terhadap Pandemi Covid-19 Pada Media Sosial Twitter," Dalam *Prisma, Prosiding Seminar Nasional Matematika*, 2021, Hlm. 1–10. Doi: 10.15294/Prisma.V6i0.45038.