

# Website Pelaporan Fasilitas Unesa Dengan Perankingan Penanganan Laporan Berbasis *Clustering K-Means* dan *Named Entity Recognition*

Algonza Dewangga Arjunantyo<sup>1</sup>, Anita Qoiriah<sup>2</sup>

<sup>1,2</sup> Jurusan Teknik Informatika, Fakultas Teknik, Universitas Negeri Surabaya

<sup>1</sup>[algonza.20050@mhs.unesa.ac.id](mailto:algonza.20050@mhs.unesa.ac.id)

<sup>2</sup>[anitaqoiriah@unesa.ac.id](mailto:anitaqoiriah@unesa.ac.id)

**Abstrak**— Masalah fasilitas muncul ketika fasilitas perlu diperbaiki atau diganti karena tidak layak pakai, seperti kebersihan gedung atau kursi meja yang rusak. Di Universitas Negeri Surabaya, berbagai fasilitas tersedia untuk dosen, mahasiswa, dan staf. Namun, karena sering digunakan atau sudah berumur, beberapa fasilitas perlu perawatan atau penggantian. Oleh karena itu, masalah fasilitas harus dilaporkan oleh civitas akademika agar dapat diperbaiki oleh pihak kampus. Seiring perkembangan waktu, jumlah laporan fasilitas dari civitas akademika meningkat dan tidak bisa diselesaikan bersamaan oleh pihak kampus, sehingga diperlukan metode untuk memprioritaskan penanganan masalah. Penelitian ini melakukan prioritas laporan masalah fasilitas menggunakan K-Means Clustering, dilanjutkan dengan ekstraksi informasi dari setiap cluster laporan menggunakan algoritma BERT dan pustaka Spacy. Penelitian ini menghasilkan sebuah website dengan fitur untuk melaporkan masalah fasilitas, mengelompokkan laporan, dan membuat perankingan penanganan menggunakan K-Means Clustering. Website ini juga mengekstrak informasi cluster menggunakan BERT dan Spacy, memungkinkan komunikasi antara admin dan pelapor, serta mengubah status laporan tunggal dan jamak dalam cluster. Pengujian silhouette coefficient untuk cluster laporan menunjukkan nilai terbaik pada cluster 5 (0.10), meskipun semua cluster termasuk dalam kategori no structure (nilai < 0.25).

**Kata Kunci**— k-means clustering, named entity recognition, BERT, Spacy, pelaporan fasilitas, website.

## I. PENDAHULUAN

Universitas adalah suatu lembaga pendidikan tinggi yang beroperasi berdasarkan kebutuhan dari civitas akademika. Untuk mendukung kepentingan civitas akademika, universitas perlu memiliki fasilitas yang mendukung agar proses kegiatan dapat berjalan dengan efisien dan tanpa hambatan. Universitas Negeri Surabaya atau UNESA tentu memiliki berbagai macam fasilitas yang memadai bagi civitas akademikanya, namun karena sering digunakan bersama atau umur fasilitas tersebut sudah cukup lama, terdapat beberapa fasilitas yang perlu dirawat kembali atau digantikan karena sudah tidak layak pakai seperti kebersihan gedung yang kurang atau kursi dan meja yang rusak. Oleh karena itu, demi kelancaran aktivitas operasional dan pendidikan, maka fasilitas tersebut perlu dilaporkan kepada pihak kampus.

Dalam era teknologi yang terus berkembang, salah satu fasilitas internet yang dapat digunakan untuk sarana penyampaian

aspirasi dan pelaporan adalah website [1]. Website adalah alat komunikasi yang menyediakan informasi untuk memperoleh layanan dan merespons masalah pengguna. Sistem pelaporan berbasis website dapat digunakan sebagai wadah bagi civitas akademika dalam melaporkan masalah fasilitas [2].

Seiring waktu, data pelaporan yang diterima oleh sistem akan semakin banyak dan pihak kampus tidak mungkin menyelesaikan laporan fasilitas secara bersamaan sehingga memerlukan suatu sistem yang menghasilkan prioritas untuk penanganan masalah fasilitas. Namun sebelum masuk ke prioritas penanganan masalah, laporan-laporan yang masuk tentu akan memiliki kemiripan satu dengan yang lainnya. Hal ini membutuhkan klasifikasi atau pengelompokkan laporan masalah fasilitas agar pada saat prioritas penanganan masalah fasilitas, tidak terdapat masalah fasilitas yang sama.

Clustering adalah salah satu teknik yang sering digunakan dalam text mining untuk mengelompokkan objek atau data ke dalam beberapa kelompok (cluster) agar data yang memiliki kesamaan akan tergabung dalam satu cluster, dengan tujuan memaksimalkan jarak antar cluster [3]. Algoritma K-Means Clustering adalah metode yang umum diterapkan dalam permasalahan clustering karena sederhana dan kemampuannya yang tinggi dalam menangani data berskala besar. Clustering data laporan adalah proses pengelompokan sejumlah dokumen laporan secara otomatis ke dalam topik-topik laporan yang spesifik berdasarkan kesamaan tema dan konten di dalamnya.

Seiring perkembangan waktu, data laporan yang dilaporkan oleh civitas akademika juga akan memuat banyak informasi mengenai detail masalah fasilitas. Namun, tidak semua informasi berkaitan dengan topik yang ingin dipilih oleh admin. Untuk memperoleh informasi rinci seperti objek laporan, pengguna harus membaca dengan teliti laporan masalah yang memberikan detail spesifik tentang objek laporan ataupun lokasi laporan dan hal ini akan memakan waktu. Oleh karena itu, proses penggalan informasi dari data laporan yang ada dapat menjadi sangat penting.

Berdasarkan permasalahan yang telah dijabarkan, maka data yang digunakan dalam penelitian ini adalah berjenis teks, sehingga metode yang dapat digunakan dalam tahap pra-pemrosesan adalah text mining dan Named Entity Recognition (NER). Text mining adalah perkembangan dari data mining yang mengubah teks yang tidak terstruktur menjadi format terstruktur guna mengidentifikasi pola yang signifikan serta mendapatkan wawasan baru [4], [5]. Named Entity

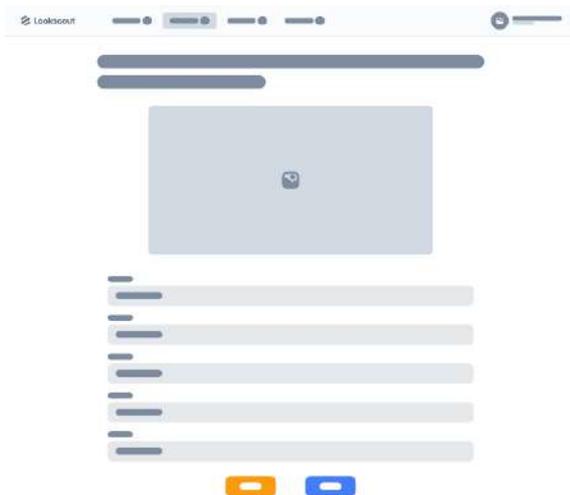




Gbr 2. Rancangan home dan fitur pelaporan fasilitas

Pada gambar 2 di halaman utama, pengguna dapat mengisi form pelaporan masalah, melihat alur pelaporan, dan melihat statistik laporan yang diajukan, ditangani, serta yang gagal.

### 2) Tampilan detail laporan



Gbr 3. rancangan detail laporan

Pada gambar 3 di page detail laporan, pengguna akan dapat melihat informasi dari laporan yang telah ia submit. Pengguna juga akan dapat melihat status laporan apakah sudah diproses atau laporan ditolak. Di dalam page ini juga pengguna akan dapat memberikan rating terhadap laporan yang terkait.

### 3) Tampilan laporan masuk, perangkingan laporan, ekstraksi informasi NER

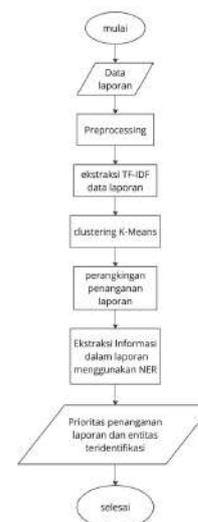
Pada gambar 4. di page perangkingan laporan, ekstraksi informasi, dll, admin akan dapat melihat informasi laporan, hasil perangkingan penanganan laporan, dan hasil ekstraksi informasi dalam tampilan tabel sehingga mempermudah dalam melihat segala informasi. Admin juga akan dapat mencari laporan tertentu di halaman laporan yang masuk.



Gbr 4. Rancangan laporan masuk, perangkingan laporan, dan NER

### D. Alur Perangkingan Penanganan Laporan dan Ekstraksi Informasi

Alur metode penelitian dimulai dari sistem menerima data laporan dari database, setelah itu data laporan akan di preprocessing untuk mengambil bagian-bagian yang penting. Setelah itu, sistem akan melakukan clustering menggunakan K-Means untuk mengelompokkan data laporan berdasarkan kesamaan topiknya [10]. Dari hasil clustering ini, sistem akan menganalisis tren menggunakan perhitungan jumlah anggota cluster. Dari informasi tren cluster ini, sistem akan menghasilkan perangkingan laporan untuk dapat diprioritaskan penanganannya.



Gbr. 5 Alur Perangkingan Penanganan Laporan dan Ekstraksi Informasi

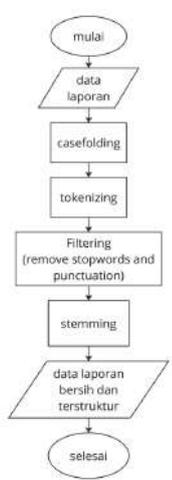
Selain proses analisis tren, Named Entity Recognition (NER) juga akan diimplementasi dalam ekstraksi informasi laporan untuk mengambil lokasi laporan setiap cluster yang sudah ada. Dengan demikian admin sistem dapat lebih mudah mengidentifikasi informasi di dalam laporan.

### E. Data Laporan

Data penelitian ini berasal dari database sistem yang akan dikembangkan. Data berupa teks laporan dengan atribut seperti deskripsi laporan. Data akan dibersihkan, diproses, dan dianalisis menggunakan *Term Frequency Inverse Document Frequency* (TF-IDF) untuk menentukan bobot kata [11], [12]. Selanjutnya, data akan dikelompokkan menggunakan K-Means Clustering. Hasil clustering akan dianalisis untuk menentukan perangkian penanganan laporan berdasarkan jumlah anggota cluster. Selain itu, informasi penting seperti objek dan lokasi laporan akan diekstraksi.

**F. Alur Preprocessing Data Laporan**

Data yang didapat dari database merupakan data mentah yang banyak mengandung *noise* dan bagian tidak informatif seperti kata tidak baku. Oleh karena itu, diperlukan preprocessing untuk mengubah data mentah ini menjadi data terstruktur yang dapat diolah oleh komputer. gambar 6 menjelaskan alur dari *preprocessing* data dalam penelitian yang akan dilakukan.



Gbr 6, alur preprocessing

**1. Casefolding**

*Case folding* adalah proses mengubah bentuk data laporan dari huruf kapital menjadi huruf kecil untuk semua kata.

Tabel 1  
 Proses Case Folding

Input	Output
Lampu di Laboratorium Kimia mengalami kerusakan. Beberapa lampu mati dan tidak menyala. Hal ini menyebabkan ruangan menjadi gelap dan mengganggu aktivitas mahasiswa dan dosen.	lampu di laboratorium kimia mengalami kerusakan. beberapa lampu mati dan tidak menyala. hal ini menyebabkan ruangan menjadi gelap dan mengganggu aktivitas mahasiswa dan dosen.

**2. Tokenizing**

*Tokenizing* adalah proses pemisahan kalimat menjadi kata per kata penyusunnya, hal ini dilakukan untuk proses selanjutnya

dan memudahkan dalam menghitung kemunculan kata dalam sebuah dokumen.

Tabel 2  
 Proses Case Folding

Input	Output
lampu di laboratorium kimia mengalami kerusakan. beberapa lampu mati dan tidak menyala. hal ini menyebabkan ruangan menjadi gelap dan mengganggu aktivitas mahasiswa dan dosen.	[ 'lampu', 'di', 'laboratorium', 'kimia', 'mengalami', 'kerusakan', '.', 'beberapa', 'lampu', 'mati', 'dan', 'tidak', 'menyala', '.', 'hal', 'ini', 'menyebabkan', 'ruangan', 'menjadi', 'gelap', 'dan', 'mengganggu', 'aktivitas', 'mahasiswa', 'dan', 'dosen', '.' ]

**3. Filtering**

Dalam tahapan ini, kata-kata umum yang tidak memiliki makna seperti "yang", "ke", "di", dan sebagainya akan dihilangkan. Selain itu, tanda baca seperti koma, titik, dan emotikon akan dihilangkan juga.

Tabel 3  
 Proses Case Folding

Input	Output
[ 'lampu', 'di', 'laboratorium', 'kimia', 'mengalami', 'kerusakan', '.', 'beberapa', 'lampu', 'mati', 'dan', 'tidak', 'menyala', '.', 'hal', 'ini', 'menyebabkan', 'ruangan', 'menjadi', 'gelap', 'dan', 'mengganggu', 'aktivitas', 'mahasiswa', 'dan', 'dosen', '.' ]	[ 'lampu', 'laboratorium', 'kimia', 'mengalami', 'kerusakan', 'lampu', 'mati', 'menyala', 'menyebabkan', 'ruangan', 'gelap', 'menjadi', 'gelap', 'dan', 'mengganggu', 'aktivitas', 'mahasiswa', 'dosen' ]

**4. Stemming**

*Stemming* adalah sebuah proses untuk menghilangkan imbuhan prefix dan suffix pada kata dalam suatu dokumen untuk diubah menjadi bentuk dasarnya.

Tabel 4  
 Proses Case Folding

Input	Output
[ 'lampu', 'laboratorium', 'kimia', 'mengalami', 'kerusakan', 'lampu', 'mati', 'menyala', 'menyebabkan', 'ruangan', 'gelap', 'mengganggu', 'aktivitas', 'mahasiswa', 'dosen' ]	[ 'lampu', 'laboratorium', 'kimia', 'alami', 'rusa', 'lampu', 'mati', 'nyala', 'sebab', 'ruang', 'gelap', 'ganggu', 'aktivitas', 'mahasiswa', 'dosen' ]

G. TF-IDF

Pada tahap ini, sistem akan menghitung bobot dari *term* dan *idf* dari dokumen dalam *database* menggunakan persamaan 1.

$$tf_{td} idf_t = tf_{td} \times idf_t \quad (1)$$

Keterangan :

$tf_{td} idf_t$  = nilai tf-idf atau bobot akhir

$tf_{td}$  = hasil perhitungan tf

$idf_t$  = hasil perhitungan idf

1. Perhitungan TF-IDF

Tabel 5  
Data Laporan

No	Data mentah	Setelah preprocessing
1	Listrik di ruang abc sering mati tiba-tiba.	listrik ruang abc mati
2	Kursi di ruang abc banyak yang patah.	kursi ruang bca patah
3	Saya menemukan beberapa kerusakan kursi di gedung bca.	temu rusak kursi gedung bca
4	Listrik sering turun di ruang cde.	listrik turun ruang cde
5	Kadang listrik padam mengganggu aktivitas belajar	listrik padam ganggu aktivitas belajar

Dari data laporan yang sudah didapat dan dilakukan pembersihan, akan dihitung skor tf-idf setiap katanya.

Tabel 6  
Contoh Perhitungan TF-IDF Pada Dokumen 1

kata	tf	idf	tf-idf
listrik	$1 + \log 4 = 1 + 0.602 = 1.602$	$\log 5 / 3 = 0.221$	$1.602 * 0.221 = 0.355$
ruang	$1 + \log 4 = 1 + 0.602 = 1.602$	$\log 5 / 2 = 0.397$	$1.602 * 0.397 = 0.637$
abc	$1 + \log 4 = 1 + 0.602 = 1.602$	$\log 5 / 1 = 0.698$	$1.602 * 0.698 = 1.119$
mati	$1 + \log 4 = 1 + 0.602 = 1.602$	$\log 5 / 1 = 0.698$	$1.602 * 0.698 = 1.119$

Perhitungan TF-IDF akan dilakukan pada dokumen 1 hingga dokumen 5.

H. Perhitungan K-Means

Tabel 7 memperlihatkan nilai tf-idf untuk semua dokumen.

Tabel 7  
TF-IDF term kata dalam dokumen

dok	listrik	ruang	abc	mati	kursi	patah	temu	rusak	gedung
1	0.36	0.64	0.11	0.11	0	0	0	0	0
2	0	0.36	0.11	0	0.64	0.11	0	0	0
3	0	0	0	0	0.67	0	0.12	0.12	0.12
4	0.36	0.64	0	0	0	0	0	0	0
5	0.38	0	0	0	0	0	0	0	0

dok	bca	sampah	tumpuk	turun	cde	padam	ganggu	aktivitas	belajar
1	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0
3	0.67	0	0	0	0	0	0	0	0
4	0	0	0	1.12	1.12	0	0	0	0
5	0	0	0	0	0	1.19	1.19	1.19	1.19

Berikut tahapan perhitungan metode K-Means :

- Menentukan jumlah *cluster* sebanyak K  
Contoh di sini menggunakan K = 2
- Penentuan pusat *cluster* (*centroid*) sejumlah K secara acak.

Tabel 8  
Contoh Pusat Cluster K = 2

cluster	listrik	ruang	abc	mati	kursi	patah	temu	rusak	gedung
1	0.36	0.64	1.12	1.12	0	0	0	0	0
2	0	0.36	1.12	0	0.64	1.12	0	0	0

cluster	bca	sampah	tumpuk	turun	cde	padam	ganggu	aktivitas	belajar
1	0	0	0	0	0	0	0	0	0

2	0	0	0	0	0	0	0	0	0
---	---	---	---	---	---	---	---	---	---

3. Menghitung jarak antara data dengan *centroid*  
Menghitung jarak antara data dan *centroid* menggunakan *euclidean distance* sesuai persamaan (2) :

$$d(x_j, c_j) = \sqrt{\sum_{j=1}^n (x_j, c_j)^2} \quad (2)$$

keterangan :

*d* : jarak

*j* : jumlah data

*x* : data

*c* : *centroid*

3.1) Jarak D1, C1

Pentuan jarak dokumen 1 dengan *centroid cluster 1* dapat dilihat pada persamaan (3)

$$\frac{\sqrt{(0.355 - 0.355)^2 + (0.637 - 0.637)^2 + (1.119 - 1.119)^2 + \dots}}{\sqrt{(1.119 - 1.119)^2 + 0 \dots + 0}} \quad (3)$$

= 0 ..... (3)

4. Membagi setiap dokumen berdasarkan kedekatannya dengan pusat *cluster* (jarak yang paling pendek).

Tabel 9

Contoh Hasil Perhitungan Jarak dalam Dokumen 1

Data	Cluster 1	Cluster 2	Cluster terdekat
Dokumen 1	0	3.344	Cluster 1

Untuk dokumen 1, karena jarak yang paling dekat atau pendek adalah *cluster 1* maka dokumen 1 masuk ke dalam *cluster 1*. Lanjutkan langkah nomor 3 dan 4 untuk setiap dokumen yang dapat dilihat pada tabel 10.

Tabel 10

Contoh hasil perhitungan centroid

Data	Cluster 1	Cluster 2	Cluster terdekat
Dokumen 2	1.77	0	cluster 2
Dokumen 3	8.39	7.3	cluster 2
Dokumen 4	1.77	0	cluster 1
Dokumen 5	8.39	7.3	cluster 1

Dari Tabel 10 didapat hasil *clustering* (k=2) yang dapat dilihat pada tabel 11.

Tabel 11

Hasil clustering seluruh dokumen

Iterasi 1	
Cluster	Dokumen
Cluster 1	dok 1, dok 4, dok 5
Cluster 2	dok 2, dok 3

Hasil *clustering* ini kemudian akan dilakukan analisis tren laporan dan dapat digunakan sebagai perangkingan penanganan laporan.

I. Perangkingan Penanganan Laporan

Perangkingan penanganan laporan adalah mengurutkan dari cluster dengan anggota paling banyak diikuti dengan cluster dengan anggota yang lebih sedikit. Perangkingan penanganan laporan dapat dilihat pada Tabel 12.

Tabel 12

Perangkingan Penanganan Laporan

Rangking	Cluster	Anggota cluster
1	1	3
2	2	2

Dari Tabel 12. *cluster* pertama akan dirangkingkan di urutan pertama disusul *cluster* dua di urutan yang kedua. Dari hasil perangkingan *clustering* ini, anggota di dalam *cluster* dapat dipakai dalam ekstraksi informasi menggunakan NER.

J. Named Entity Recognition

Named Entity Recognition (NER) adalah proses mengekstrak informasi dan memproses dokumen secara terstruktur dengan fokus pada identifikasi nama orang, lokasi, organisasi, dan waktu [6], [7]. Dalam penelitian ini, identifikasi entitas akan berfokus pada objek atau subjek laporan, serta lokasi laporan. Data yang dipakai untuk proses ekstraksi informasi yaitu seluruh data laporan di dalam cluster. Data yang dipakai ini dapat dilihat pada tabel 13.

Tabel 13.

Contoh data untuk ekstraksi informasi NER

Cluster	Dokumen
---------	---------

Cluster 1	<ul style="list-style-type: none"> <li>listrik ruang abc mati</li> <li>listrik turun ruang cde</li> <li>listrik padam ganggu aktivitas belajar</li> </ul>
Cluster 2	<ul style="list-style-type: none"> <li>kursi ruang bca patah</li> <li>temu rusak kursi gedung bca</li> </ul>

Proses ekstraksi informasi terdapat 3 metode yaitu :

1. BERT-BASE

Tabel 14  
Contoh hasil NER BERT Cluster 1

Kalimat	Entitas
Listrik padam ganggu aktivitas belajar. Listrik ruang abc mati. Listrik turun ruang cde.	<ul style="list-style-type: none"> <li>· Abc (location)</li> <li>· Cde (location)</li> </ul>

2. Pustaka Spacy

Tabel 15  
Contoh hasil NER spaCy cluster 1

Entitas	Kategori
Listrik ruang	person
Abc	organization
Listrik turun ruang	person
Listrik padam ganggu aktivitas	person

3. NER Pustaka spaCy

Tabel 16  
Contoh hasil NER spaCy cluster 1

Entitas	Kategori
ABC	Organization
CDE	Organization

K. Silhouette Coefficient

Silhouette coefficient adalah metode untuk mengukur kekuatan dan kualitas suatu cluster dengan konsep cohesion dan separation [3]. Cohesion menghitung jarak antar data dalam satu cluster, sementara separation menghitung jarak antar cluster. Berikut adalah tahapan evaluasi cluster menggunakan silhouette coefficient. :

1. Mengukur rata-rata jarak antara data dan seluruh data lain dalam suatu cluster dengan menggunakan persamaan (4) sebagai berikut.

$$a(i) = \frac{1}{[A]-1} \sum_{j \in A, j \neq i} d(i, j) \quad (4)$$

keterangan :

$a(i)$  : Nilai rata-rata antara jarak data dengan jarak setiap data dalam suatu cluster

$[A]$  : jumlah anggota cluster

$d(i, j)$  : jarak data dengan jarak setiap data di dalam cluster

2. Mengukur rata-rata jarak antara data dan seluruh data lain dalam cluster lain dengan menggunakan persamaan (5) sebagai berikut.

$$b(i) = \frac{1}{[A]} \sum_{j \in C} d(i, j) \quad (5)$$

keterangan :

$b(i)$  : Nilai rata-rata antara jarak data dengan jarak setiap data dalam cluster lain

$[A]$  : jumlah anggota cluster

$d(i, j)$  : jarak data dengan jarak setiap data dalam cluster lain

3. Mengukur silhouette coefficient menggunakan persamaan (6) sebagai berikut.

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (6)$$

keterangan :

$s(i)$  : Nilai silhouette coefficient

$a(i)$  : Nilai rata-rata antara jarak data dengan jarak setiap data dalam suatu cluster

$b(i)$  : Nilai rata-rata antara jarak data dengan jarak setiap data dalam cluster lain

$\max(a(i), b(i))$  : Nilai terbesar di antara  $a(i)$  dan  $b(i)$

Silhouette coefficient memiliki rentang nilai dari 1 hingga -1. Nilai dianggap baik jika positif, dengan  $a(i) < b(i)$  dan  $a(i)$  mendekati 0, yang menghasilkan nilai mendekati 1. Jika  $s(i) = 1$ , data i berada di cluster yang benar; jika  $s(i) = 0$ , data i tidak jelas posisinya; dan jika  $s(i) = -1$ , data i berada di cluster yang salah. Silhouette coefficient mengukur seberapa rapat data terkelompok dalam suatu cluster. Metrik ini diklasifikasikan ke dalam beberapa kategori sebagai berikut.

$$\begin{aligned}
 0.7 < s(i) < 1, & \text{Strong Structure} \\
 0.5 < s(i) < 0.7, & \text{Medium structure} \\
 0.25 < s(i) < 0.5, & \text{Weak Structure} \\
 s(i) < 0.25, & \text{No Structure}
 \end{aligned}$$

L. Pengujian Blackbox

Pengujian black box dilakukan dengan memberikan input ke aplikasi dan memeriksa outputnya tanpa perlu memahami kode program. Hasil pengujian didokumentasikan dalam tabel yang mencakup deskripsi pengujian, fitur yang diuji, hasil yang diharapkan, hasil aktual, dan kesimpulan. Tabel 17. menunjukkan rancangan pengujian black box dari sisi pengguna, sedangkan Tabel 18. dari sisi admin.

Tabel 17  
Tabel rancangan pengujian black box pengguna

Fitur yang diujikan	Hasil yang diharapkan	Hasil yang didapatkan	keterangan
Button submit form laporan fasilitas	Button form laporan fasilitas berfungsi dengan baik		
Fitur lihat status laporan	Pengguna dapat melihat status laporan apakah sudah ada perbaikan atau belum		
Fitur chat dengan admin	Pengguna dapat melakukan chat dengan admin di halaman detail laporan dari laporan yang telah dibuat		
Fitur rating penanganan laporan	Pengguna dapat memberikan rating terhadap penanganan laporan fasilitas		

Tabel 18  
Tabel rancangan pengujian black box admin

Fitur yang diujikan	Hasil yang diharapkan	Hasil yang didapatkan	keterangan
Fitur perangkingan laporan fasilitas	administrator dapat melihat perangkingan penanganan laporan fasilitas		
Fitur melihat laporan	administrator dapat melihat semua laporan yang masuk dan detailnya		
Fitur chat dengan pengguna	administrator dapat melakukan chat dengan pengguna		
Fitur ekstraksi informasi penting laporan	administrator dapat melihat informasi penting pada setiap cluster laporan		
Fitur memperbarui status laporan	administrator dapat merubah status laporan dari belum ada penanganan, menjadi sudah ditangani		

Bab ini menyajikan hasil dari penelitian “Website Pelaporan Fasilitas UNESA Dengan Perangkingan Penanganan Laporan Berbasis Clustering K-Means Dan Named Entity Recognition”. Selain itu, hasil-hasil tersebut akan dibahas secara mendalam untuk memberikan pemahaman yang lebih menyeluruh.

A. Hasil Website

1) Pelaporan Fasilitas

Fitur Pelaporan masalah fasilitas pada gambar 7 digunakan oleh pengguna yang telah login ke dalam website. Fitur ini memungkinkan pengguna untuk melaporkan masalah fasilitas kepada admin dengan mengisi form yang tersedia. Form pelaporan fasilitas mencakup input untuk judul laporan, lokasi, deskripsi laporan, dan gambar laporan, yang terlihat pada gambar 7.

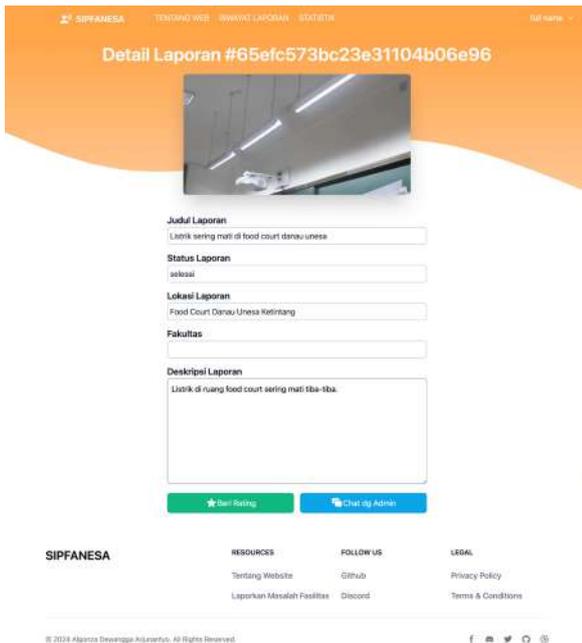


Gbr 7. tampilan home dan pelaporan fasilitas

2) Detail Laporan

Fitur detail laporan pada gambar 8 digunakan oleh pengguna untuk meninjau informasi yang lebih mendalam mengenai laporan yang telah pengguna unggah, termasuk status dari penanganan laporan tersebut. Di dalam fitur ini juga memungkinkan pengguna untuk memberikan rating terhadap proses penanganan laporan yang pengguna ajukan.

III. HASIL DAN PEMBAHASAN



Gbr 8. detail laporan

3) Perangkingan Penanganan Laporan

Fitur Perangkingan Laporan

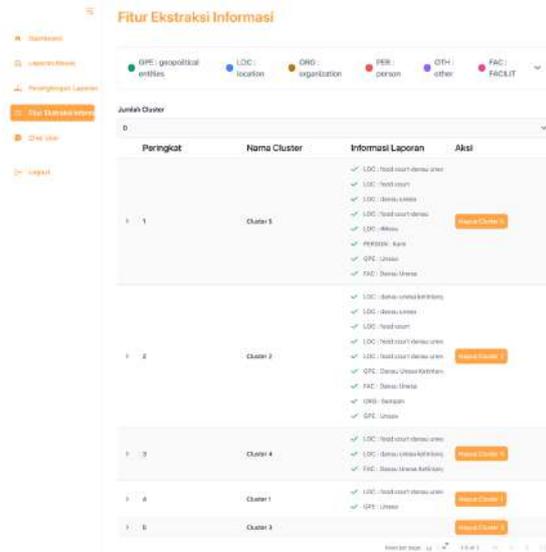


Gbr 9. perangkingan penanganan laporan

Fitur perangkingan penanganan laporan pada gambar 9 digunakan oleh admin untuk melihat prioritas penanganan laporan. Fitur ini mengelompokkan laporan ke dalam cluster menggunakan K-Means Clustering. Fitur ini akan memastikan suatu cluster yang berisi jumlah anggota yang paling banyak, akan urutan paling atas disusul dengan cluster dengan jumlah anggota lebih sedikit.

4) Ekstraksi Informasi

Fitur ekstraksi informasi digunakan oleh admin untuk mengekstrak informasi di dalam laporan seperti objek laporan, subjek laporan, atau lokasi dari laporan yang dapat dilihat pada gambar 10, Fitur ini menggunakan Named Entity Recognition (NER) berbasis BERT serta pustaka spacy.



Gbr 10. ekstraksi informasi laporan

B. Pembahasan

1) Membuat server

Hal pertama yang dilakukan peneliti dalam membangun website ini adalah membuat server berbasis python untuk melakukan perangkingan laporan serta tugas lainnya. Untuk mempermudah proses pengembangan, peneliti menggunakan framework *Fastapi*. Sebelum menginstall fastapi, dalam sistem operasi harus sudah menginstall python3 terlebih dahulu, karena peneliti menggunakan sistem operasi macOS, maka peneliti dapat menginstall python melalui homebrew.



Gbr 11. menginstall homebrew

Setelah menginstall homebrew, peneliti dapat langsung menginstall python3 dengan perintah pada gambar 12 berikut.



Gbr 12. menginstall python

Setelah terdapat python3 dalam sistem operasi peneliti, maka langkah selanjutnya yaitu melakukan instalasi Fastapi dengan perintah pada gambar 13 sebagai berikut.

```

instalasi fatapi
pip install fastapi "uvicorn[standard]"
    
```

Gbr 13. menginstall *fastapi*

Langkah selanjutnya adalah membuat *file main.py* sebagai *root server python*, dan memasukkan kode pada gambar 14 berikut.

```

file root server
from fastapi import FastAPI
app = FastAPI()
@app.get("/")
def read_root():
    return {"Hello": "World"}
    
```

Gbr 14. file root server

Setelah selesai membuat *file root server*, peneliti menjalankan server *python* dengan perintah pada gambar 15 di terminal. Setelah server berjalan, pengguna sudah bisa mengonsumsi *API* dari aplikasi *server python* yang sudah peneliti buat.

```

menjalankan server
uvicorn main:app --reload
    
```

Gbr 15. Menjalankan server python

2) Membuat Fungsi Perangkingan Penanganan Laporan  
Fungsi perangkingan laporan menggunakan teks yang sudah di pre-processing, TF-IDF serta K-Means dalam mengelompokkan laporan ke dalam klasternya masing-masing. Fungsi ini mempunyai parameter data laporan dari database dan jumlah klaster yang ingin diciptakan oleh admin website yang dapat dilihat pada gambar 16. Langkah pertama dalam perangkingan laporan yaitu pre-processing teks dengan memecah kata per kata di dalam kalimat serta mengubah huruf kapital menjadi huruf kecil per kata. Lalu menghilangkan tanda baca dan stopwords atau kata-kata yang kurang bermakna seperti ke, di, dll. Selanjutnya adalah perhitungan TF-IDF untuk dokumen laporan yang sudah di pre-processing untuk kemudian di clustering sehingga setiap laporan masuk ke dalam cluster yang mirip dengan dirinya sendiri. Langkah terakhir adalah melakukan perangkingan cluster berdasarkan jumlah anggota cluster. Hasil perangkingan ini akan menjadi acuan perangkingan penanganan laporan.

```

def perform_clustering_Bev(laporan, jumlah_cluster):
    stop_words = set(stopwords.words("indonesian"))
    sentences = [laporan["deskripsi"] for laporan in laporan]
    tokenize_words = [word_tokenize(cleaning(word)) for word in sentences]
    filtered_words = [
        word for word in tokenize_word if word.lower() not in stop_words
        for tokenize_word in tokenize_words
    ]
    # filtered_words = tokenize_words
    clean_words = [" ".join(sentence) for sentence in filtered_words]
    not_clean_words = [" ".join(sentence) for sentence in tokenize_words]
    # print(clean_words)
    tfidf_vectorizer = TfidfVectorizer()
    tfidf_matrix = tfidf_vectorizer.fit_transform(clean_words)
    tfidf_df = pd.DataFrame(tfidf_matrix.toarray())
    # Melakukan clustering dengan K-Means
    num_clusters = jumlah_cluster # Ganti dengan jumlah cluster yang diinginkan
    kmeans = KMeans(n_clusters=num_clusters)
    kmeans.fit(tfidf_matrix)
    
```

Gbr 16. potongan kode fungsi perangkingan laporan

3) Pembahasan Perangkingan Laporan (K-MEANS)  
Misal terdapat tujuh buah data laporan pada tabel 19 sebagai berikut.

Tabel 19  
contoh laporan yang masuk

No	Judul	Deskripsi	Fakultas	Jurusan	Lokasi
1	kursi food court sering patah	Kursi di food court danau unesa sering patah.	-	-	Food Court Danau Unesa Ketintang
2	kursi di food court	Kursi di food court danau unesa sudah sangat usang.	-	-	Food court Danau Unesa Ketintang
3	Tempat duduk tidak nyaman di food court danau unesa	tempat duduk di food court danau unesa tidak nyaman.	-	-	Food Court Danau Unesa Ketintang
4	Kursi food court perlu diperbaiki	Kursi di food court perlu diperbaiki.	-	-	Food Court Danau Unesa Ketintang
5	tempat sampah penuh	Saya menemukan tempat sampah yang penuh di food court.	-	-	Food Court Danau Unesa Ketintang
6	Banyak sampah di sekitar danau unesa	Saya menemukan banyak sampah di sekitar danau unesa.	-	-	Food Court Danau Unesa Ketintang
7	tempat sampah di food court	Tempat sampah di food court danau unesa penuh.	-	-	Food Court Danau Unesa Ketintang

Langkah pertama yang dilakukan adalah mengambil data dari kolom deskripsi pelaporan lalu melakukan tokenisasi (memecah setiap kata) serta membersihkan setiap laporan dari kata-kata umum yang tidak memiliki makna seperti "yang", "ke", "di", dan sebagainya akan dihilangkan. Selain itu, tanda baca seperti koma, titik, dan emotikon akan dihilangkan juga. Hasilnya dapat dilihat pada tabel 20.

Tabel 20  
Hasil tokenisasi dan cleaning

No	Deskripsi laporan	Hasil
1	Kursi di food court danau unesa sering patah.	["kursi","di","food","court","danau","unesa","sering","patah"]
2	Kursi di food court danau unesa sudah sangat usang.	["kursi","di","food","court","danau","unesa","sudah","sangat","usang"]
3	tempat duduk di food court danau unesa tidak nyaman.	["tempat","duduk","di","food","court","danau","unesa","tidak","nyaman"]
4	Kursi di food court perlu diperbaiki.	["kursi","di","food","court","perlu","diperbaiki"]
5	Saya menemukan tempat sampah yang penuh di food court.	["saya","menemukan","tempat","sampah","yang","penuh","di","food","court"]
6	Saya menemukan banyak sampah di sekitar danau unesa.	["saya","menemukan","banyak","sampah","di","sekitar","danau","unesa"]
7	Tempat sampah di food court danau unesa penuh.	["tempat","sampah","di","food","court","danau","unesa","penuh"]

Langkah selanjutnya yang dilakukan adalah menghilangkan stopwords di dalam array dari data laporan, stopwords adalah kata-kata umum yang tidak memiliki makna seperti "yang", "ke", "di", dan sebagainya. Hasilnya dapat dilihat pada table XXI.

Tabel 21  
Menghilangkan stopwords dari data laporan

no	deskripsi laporan	hasil
1	Kursi di food court danau unesa sering patah.	["kursi","food","court","danau","unesa","patah"]
2	Kursi di food court danau unesa sudah sangat usang.	["kursi","food","court","danau","unesa","usang"]
3	tempat duduk di food court danau unesa tidak nyaman.	["duduk","food","court","danau","usang","unesa","nyaman"]
4	Kursi di food court perlu diperbaiki.	["kursi","food","court","diperbaiki"]

5	Saya menemukan tempat sampah yang penuh di food court.	["menemukan","sampah","penuh","food","court"]
6	Saya menemukan banyak sampah di sekitar danau unesa.	["menemukan","sampah","danau","unesa"]
7	Tempat sampah di food court danau unesa penuh.	["sampah","food","court","danau","unesa","penuh"]

Langkah selanjutnya adalah menggabungkan kata yang sudah dipisah kembali menjadi satu kalimat utuh lalu menghitung tf-idf dari setiap laporan yang ada. Tabel 22 memperlihatkan hasilnya.

Tabel 22  
TF-IDF kata dalam dokumen

dok	court	danau	diperbaiki	duduk	food	kursi	menemukan
1	0.3	0.34	0	0	0.3	0.45	0
2	0	0	0	0	0.3	0.45	0
3	0.27	0	0	0.57	0	0	0
4	0.34	0	0.72	0	0.34	0.51	0
5	0.31	0	0	0	0.31	0	0.54
6	0	0.41	0	0	0	0	0.62
7	0.32	0.36	0	0	0.32	0	0

dok	nyaman	patah	penuh	sampah	unesa	usang
1	0	0.63	0	0	0.34	0
2	0	0	0	0	0.34	0.63
3	0.57	0	0	0	0.31	0
4	0	0	0	0	0	0
5	0	0	0.54	0.46	0	0
6	0	0	0	0.53	0.41	0
7	0	0	0.56	0.48	0.36	0

Tabel 23  
Cluster data laporan k = 2

	court	danau	diperbaiki	duduk	food	kursi	menemukan
1	0.3	0.34	0	0	0.3	0.45	0
2	0.31	0	0	0	0.31	0	0.54

	nyaman	patah	penuh	sampah	unesa	usang
1	0	0.63	0	0	0.34	0

2	0	0	1	0	0	0
---	---	---	---	---	---	---

Langkah selanjutnya adalah mencari jarak antara dokumen satu hingga tujuh dengan pusat klaster dengan persamaan 7 sebagai berikut.

$$d(x_j, c_j) = \sqrt{\sum_{j=1}^n (x_j, c_j)^2} \quad (7)$$

keterangan :

d : jarak

j : jumlah data

x : data

c : centroid

Langkah selanjutnya adalah penentuan jarak dokumen 1 dengan centroid cluster 1 dapat dilihat pada persamaan (8)

$$\sqrt{(0,2982235473028 - 0,2727260061277975)^2 + \dots} = 0 \dots \dots \dots (8)$$

Penentuan jarak dokumen 1 dengan centroid cluster 2 dapat dilihat pada persamaan (9)

$$\sqrt{(0,2982235473028603 - 0,311007383779932)^2 + \dots} = 1,119 \dots \dots \dots (9)$$

1. Membagi setiap dokumen berdasarkan kedekatannya dengan pusat cluster (jarak yang paling pendek).

Tabel 24.

Contoh Hasil Perhitungan Jarak dalam Dokumen 1

Data	Cluster 1	Cluster 2	Cluster terdekat
Dokumen 1	0	1.261	Cluster 1

Untuk dokumen 1, karena jarak yang paling dekat atau pendek adalah cluster 1 maka dokumen 1 masuk ke dalam cluster 1. Lanjutkan langkah nomor 3 dan 4 untuk setiap dokumen yang dapat dilihat pada tabel 25.

Tabel 25

Contoh Hasil Perhitungan Jarak Centroid

Data	Cluster 1	Cluster 2	Cluster terdekat
Dokumen 2	0.89	1.28	Cluster 1
Dokumen 3	1	1	Cluster 1
Dokumen 4	1	1	Cluster 1
Dokumen 5	1	0	Cluster 2
Dokumen 6	1	0.91	Cluster 2
Dokumen 7	1	0.75	Cluster 2

Dari Tabel 25 didapat hasil clustering (k=2) yang dapat dilihat pada tabel 26.

Tabel 26

Hasil clustering seluruh dokumen

Iterasi 1	
Cluster	Dokumen
Cluster 1	dok 1, dok 2, dok 3, dok 4
Cluster 2	dok 5, dok 6, dok 7

Langkah selanjutnya adalah melakukan perangkingan dengan menyortir cluster dengan jumlah anggota yang paling banyak sehingga mendapat perangkingan pada tabel 27 sebagai berikut.

Tabel 27

Hasil perangkingan laporan

Rangking	Cluster
1	cluster 1
2	cluster 2

4) Membuat Fungsi Ekstraksi Informasi  
Fungsi ekstraksi informasi memudahkan admin untuk mengidentifikasi nama, lokasi, organisasi, waktu, atau objek dalam cluster laporan. Menggunakan algoritma BERT dan pustaka spacy, fungsi ini memberikan hasil ekstraksi yang bervariasi. Dengan BERT, peneliti memanfaatkan API dari huggingface dan model pre-trained “cahya/ bert-base-indonesian-NER” untuk menghasilkan entitas dalam laporan. Hasilnya adalah entitas teridentifikasi beserta jenisnya, seperti tempat, nama orang, atau waktu.

```

1 def ekstraksi_ner_trans(sentences):
2     translator = Translator()
3     translation = translator.translate(sentences, dest="en", src="id")
4     # print(translation: {translation.text})
5     after_translation = translation.text
6
7     # Load the spacy engine:
8     nlp = spacy.load("en_core_web_sm")
9
10    doc = nlp(after_translation)
11
12    def translate_to_indonesian(text):
13        translation = translator.translate(text, dest="id", src="en")
14        return translation.text
15
16    # Menjumlahkan setiap entitas secara terpisah dan membangun list of dictionaries
17    unique_translations = set()
18    translations = []
19    for ent in doc.ents:
20        # If ent.start < len(doc) and ent.end < len(doc):
21        #     translated_text = translate_to_indonesian(ent.text)
22        #     translations.append({"text": str(translated_text), "label": ent.label_})
23        if ent.start < len(doc) and ent.end < len(doc):
24            translated_text = translate_to_indonesian(ent.text)
25            if translated_text not in unique_translations:
26                unique_translations.add(translated_text)
27                translations.append({"text": str(translated_text), "label": ent.label_})
28
29    return {"entities": translations}

```

Gbr 17. Fungsi Ekstraksi Informasi Pustaka Spacy

Untuk spacy, peneliti menggunakan pustaka terjemahan dari bahasa Indonesia ke bahasa Inggris terlebih dahulu. Setelah menerjemahkan laporan, spacy digunakan untuk mengekstrak informasi dari kalimat yang diterjemahkan [8], [9]. Hasil ekstraksi kemudian diterjemahkan kembali ke bahasa Indonesia agar mudah dipahami oleh admin.

5) Pembahasan ekstraksi informasi NER  
Misal terdapat sebuah kluster laporan dalam tabel 28 sebagai berikut.

Tabel 28  
contoh laporan yang masuk

no	judul	deskripsi	fakultas	jurusan	lokasi
1	kursi food court sering patah	Kursi di food court danau unesa sering patah.	-	-	Food Court Danau Unesa Ketintang
2	kursi di food court	Kursi di food court danau unesa sudah sangat usang.	-	-	Food court Danau Unesa Ketintang
3	Tempat duduk tidak nyaman di food court danau unesa	tempat duduk di food court danau unesa tidak nyaman.	-	-	Food Court Danau Unesa Ketintang
4	Kursi food court perlu diperbaiki	Kursi di food court perlu diperbaiki.	-	-	Food Court Danau Unesa Ketintang

Langkah pertama yang dilakukan adalah mengambil kolom deskripsi lalu menggabungkan seluruh kalimat menjadi satu buah kalimat atau satu buah paragraf seperti pada tabel 29 berikut.

Tabel 29  
Gabungan teks laporan

No	Gabungan deskripsi laporan
1	Kursi di food court danau unesa sering patah. Kursi di food court danau unesa sudah sangat usang. tempat duduk di food court danau unesa tidak nyaman. Kursi di food court perlu diperbaiki.

Selanjutnya gabungan teks ini akan dimasukkan ke dalam fungsi NER dimana fungsi tersebut akan mengekstrak informasi. cara kerja NER ini adalah menyesuaikan dataset yang sudah pernah dilatih sebelumnya [8], misalnya terdapat dataset yang sudah pernah diberi *tagging person* untuk “udin”, *location* untuk “danau” dan juga setelah kata “di” adalah sebuah *location*, maka pada saat fungsi NER ini dijalankan akan menghasilkan sesuai dengan *tagging* yang pernah dilatih. Tabel 30 memperlihatkan hasil dari ekstraksi informasinya.

Tabel 30  
Hasil Ekstraksi Informasi

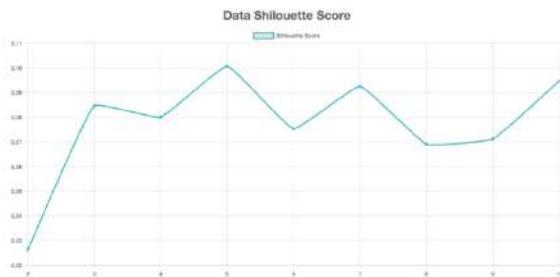
nama tagging	hasil
location	food court danau unesa
geopolitical entities	unesa

C. Pengujian

1) Pengujian silhouette coefficient

Silhouette coefficient adalah metode untuk mengukur kekuatan dan kualitas cluster dengan konsep cohesion (jarak antar data dalam cluster) dan separation (jarak antar cluster) [3]. Nilainya berkisar antara -1 hingga 1, dengan nilai positif menunjukkan cluster yang baik. Silhouette coefficient menunjukkan seberapa rapat data terkelompok dalam suatu cluster. Hasilnya adalah cluster 2 dengan nilai 0.027, cluster 3 dengan nilai 0.085, cluster 4 dengan nilai 0.08, cluster 5 dengan nilai 0.10, cluster 6 dengan nilai 0.075, cluster 7 dengan nilai 0.092, cluster 8 dengan nilai 0.069, cluster 9 dengan nilai 0.071, dan cluster 10 dengan nilai 0.095. Ini menunjukkan bahwa semua cluster memiliki structure yang buruk (nilai < 0.25), kecuali cluster 5 yang memiliki performa terbaik.

Fitur Silhouette Score



Gbr 18. Grafik Skor Silhouette Coefficient

2) *pengujian blackbox*

*Pengujian black box* adalah pengujian yang dilakukan dengan cara memberikan input ke aplikasi dan memeriksa outputnya. Pada pengujian ini, fokus utama adalah memastikan bahwa semua fitur aplikasi berjalan sesuai dengan spesifikasi dan kebutuhan yang telah ditetapkan. Berdasarkan hasil pengujian blackbox yang telah dilakukan, seluruh fungsi dan fitur aplikasi berhasil diuji dan menunjukkan kinerja yang sesuai dengan harapan. Tidak ditemukan kesalahan atau ketidaksesuaian antara output yang dihasilkan dengan ekspektasi, sehingga dapat disimpulkan bahwa aplikasi memenuhi standar kualitas yang diinginkan. Tabel pengujian blackbox beserta hasilnya dapat dilihat pada tabel 31 sampai 32.

Tabel 31

Tabel rancangan pengujian black box pengguna

Fitur yang diujikan	Hasil yang diharapkan	Hasil yang didapatkan	keterangan
Button submit form laporan fasilitas	Button form laporan fasilitas berfungsi dengan baik	Pengguna berhasil melakukan submit form laporan	Berhasil
Fitur lihat status laporan	Pengguna dapat melihat status laporan apakah sudah ada perbaikan atau belum	Pengguna dapat melihat status laporan apakah sudah ada perbaikan atau belum	Berhasil
Fitur chat dengan admin	Pengguna dapat melakukan chat dengan admin di halaman detail laporan dari laporan yang telah dibuat	Pengguna dapat melakukan chat dengan admin di halaman detail laporan dari laporan yang telah dibuat	Berhasil
Fitur rating penanganan laporan	Pengguna dapat memberikan	Pengguna dapat memberikan	Berhasil

	rating terhadap penanganan laporan fasilitas	rating terhadap penanganan laporan fasilitas	
--	--	--	--

Tabel 32

Tabel rancangan pengujian black box admin

Fitur yang diujikan	Hasil yang diharapkan	Hasil yang didapatkan	keterangan
Fitur perangkangan laporan fasilitas	administrator dapat melihat perangkangan penanganan laporan fasilitas	administrator dapat melihat perangkangan penanganan laporan fasilitas	Berhasil
Fitur melihat laporan	administrator dapat melihat semua laporan yang masuk dan detailnya	administrator dapat melihat semua laporan yang masuk dan detailnya	Berhasil
Fitur chat dengan pengguna	administrator dapat melakukan chat dengan pengguna	administrator dapat melakukan chat dengan pengguna	Berhasil
Fitur ekstraksi informasi penting laporan	administrator dapat melihat informasi penting pada setiap cluster laporan	administrator dapat melihat informasi penting pada setiap cluster laporan	Berhasil
Fitur memperbarui status laporan	administrator dapat merubah status laporan dari belum ada penanganan, menjadi	administrator dapat merubah status laporan dari belum ada penanganan	Berhasil

	sudah ditangani	an, menjadi sudah ditangani	
--	-----------------	-----------------------------	--

#### IV. KESIMPULAN

Beberapa kesimpulan yang dapat diambil dari penelitian ini adalah sebagai berikut :

- 1) *Website* Pelaporan Fasilitas Unesa adalah *website* yang mempunyai fitur untuk melaporkan masalah fasilitas, mengelompokkan dokumen laporan menggunakan *K-Means Clustering*, membuat perangkian penanganan laporan, mengekstrak informasi menggunakan NER, komunikasi antara admin dan pelapor, merubah status laporan tunggal, merubah status laporan dalam *cluster*, dan semua fitur ini sudah berjalan dengan baik.
- 2) Untuk menghasilkan perangkian penanganan laporan yaitu menggunakan *K-Means Clustering* untuk menghasilkan *cluster* laporan. Dari hasil *cluster* ini, kemudian dihitung jumlah anggota setiap *cluster* laporan, jumlah anggota ini akan diurutkan dari yang terbanyak akan dirangkikan paling atas disusul dengan *cluster* dengan jumlah anggota lebih sedikit.
- 3) Untuk menghasilkan ekstraksi informasi yaitu menggunakan algoritma BERT dan pustaka spacy pada gabungan laporan setiap *cluster* yang sudah dihasilkan oleh *K-Means Clustering* sebelumnya.
- 4) Nilai silhouette coefficient dari *cluster* dua hingga sepuluh berturut-turut yaitu 0.027, 0.085, 0.08, 0.10, 0.075, 0.092, 0.069, 0.071, 0.095. Hal ini menunjukkan structure dari *cluster* dua hingga sepuluh yakni no structure karena bernilai < 0.25 serta *cluster* berjumlah lima memiliki performa atau akurasi yang terbaik diantara *cluster* lainnya.

#### V. SARAN

Berdasarkan hasil penelitian dan pembahasan, maka dapat disusun beberapa saran penelitian, sebagai berikut.

- 1) Edukasi dan promosi kepada civitas akademik di UNESA tentang *website* ini yang berfungsi menampung laporan masalah fasilitas. Kampanye edukatif untuk menarik pengguna serta investor.
- 2) Mengembangkan fitur baru serta desain antarmuka yang lebih menarik dan interaktif kepada pengguna di masa yang akan datang untuk memaksimalkan pengalaman pengguna.
- 3) Terus memberikan pemeliharaan *website*, pemantauan kinerja, agar tetap stabil dan tidak mengalami gangguan dikarenakan *website* ini berpotensi digunakan oleh banyak orang.

#### UCAPAN TERIMA KASIH

Puji syukur kepada Tuhan Yang Maha Esa atas rahmat dan berkatnya penulis dapat menyelesaikan artikel ilmiah ini. Terimakasih kepada orang tua yang selalu memberikan dukungan moral dan materil selama pengerjaan artikel ilmiah ini. Penulis juga mengucapkan banyak terimakasih kepada pembimbing skripsi atas saran dan masukan kepada penulis. Teman-teman penulis yang sudah menemani dan membantu selama proses pengerjaan artikel ilmiah ini.

#### REFERENSI

- [1] Shidiqi, Harum Mu'alimus, Abdillah Baradja, and Tri Irianto Tjendrowasono. "Pembuatan Aplikasi Mobile Helpdesk Ticketing System dalam Penyampaian Aspirasi pada Perguruan Tinggi." *Surakarta Informatik Journal* 5.2 (2023): 9-16.
- [2] Islamiyah, Nabila Aydinda. "Aplikasi Kepuasan Masyarakat Pelayanan Desa Rejoagung Berbasis Web dan Android." *Jurnal Informatika dan Multimedia* 15.1 (2023): 6-13.
- [3] Irianto, Maulana Rafael, Achmad Maududie, and Fajrin Nurman Arifin. "Implementation of K-Means Clustering Method for Trend Analysis of Thesis Topics (Case Study: Faculty of Computer Science, University of Jember)." *BERKALA SAINSTEK* 10.4 (2022): 210-226.
- [4] Silalahi, Natalia, and Guidio Leonarde Ginting. "Rekomendasi Berita Berkaitan dengan Menerapkan Algoritma Text Mining dan TF-IDF." *Bulletin of Computer Science Research* 3.4 (2023): 276-282.
- [5] Dharmendra, I. K., Kusuma, I. G. N. A., Dewi, I. A. M. C., Edwar, "Implementasi Text Mining Untuk Klasifikasi Opini Alumni Pada Perguruan Tinggi". *Jurnal Teknologi Informasi dan Komputer* 9.3.
- [6] Hidayatullah, Ahmad Fathan, et al. "Named Entity Recognition on Tourist Destinations Reviews in the Indonesian Language." *Jurnal Linguistik Komputasional* 6.1 (2023): 30-35.
- [7] Irfan, Muhammad. "Named Entity Recognition Untuk Data Review Tempat Wisata Dengan Metode "Bidirectional Encoder Representations From Transformers". Tesis. Universitas Islam Indonesia. (2021).
- [8] Chantrapornchai, Chantana, and Aphisit Tunsakul. "Information extraction tasks based on BERT and SpaCy on tourism domain." *ECTI Transactions on Computer and Information Technology (ECTI-CIT)* 15.1 (2021): 108-122.
- [9] Kumar, Murari, et al. "An Algorithm for Automatic Text Annotation for Named Entity Recognition using spaCy Framework." Dissertation. Indian Agricultural Research Institute (2023).
- [10] Utami, N. W., & Eka Putra, I. G. J. (2022). Text Minig Clustering Untuk Pengelompokan Topik Dokumen Penelitian Menggunakan Algoritma K-Means Dengan Cosine SimilarITY. *Jurnal Informatika Teknologi dan Sains*, 4(3), 255–259.
- [11] Widaningrum, Ida, et al. "Algoritma Term Frequency-Inverse Document Frequency (TF-IDF) dan K-Means Clustering Untuk Menentukan Kategori Dokumen." *Prosiding SISFOTEK* 6.1 (2022): 145-149.
- [12] Darwis, Muhammad, et al. "Implementation of TF-IDF Algorithm and K-mean Clustering Method to Predict Words or Topics on Twitter." *JISA (Jurnal Informatika dan Sains)* 3.2 (2020): 49-55.