

## Penerapan Algoritma *K-Means* dan *Multi-Class Support Vector Machine* dalam Identifikasi Tingkat Kesadaran Keamanan Informasi Pegawai Pemerintah Kota Kendari

Fiqram Faathir<sup>1\*</sup>, Isnawaty<sup>1</sup>, Rizal Adi Saputra<sup>1</sup>

<sup>1</sup>Program Studi Teknik Informatika, Universitas Halu Oleo, Indonesia.

---

### Artikel Info

#### Kata Kunci:

HAIS-Q;  
KAB;  
Keamanan Informasi;  
Machine Learning;  
Pegawai Pemerintah.

#### Keywords:

HAIS-Q;  
KAB;  
Information Security;  
Machine Learning;  
Government Employees.

---

#### Riwayat Artikel:

Submitted: 22 Mei 2026  
Accepted: 18 Agustus 2026  
Published: 21 Agustus 2026

**Abstrak:** Perkembangan teknologi informasi meningkatkan risiko keamanan data, terutama pada instansi pemerintahan yang mengelola informasi strategis. Penelitian ini bertujuan menerapkan metode *machine learning* untuk mengidentifikasi tingkat kesadaran keamanan informasi pegawai Pemerintah Kota Kendari menggunakan pendekatan *Human Aspects of Information Security Questionnaire* (HAIS-Q). Algoritma *K-Means* digunakan untuk membentuk kategori tingkat kesadaran berdasarkan karakteristik data, sedangkan *Multi-Class Support Vector Machine* (SVM) digunakan untuk membangun model klasifikasi. Evaluasi kluster menggunakan *Elbow Method*, *Silhouette Score*, dan *Davies-Bouldin Index* (DBI) menghasilkan tiga kluster optimal dengan nilai *Silhouette Score* sebesar 0,428 dan DBI sebesar 0,769. Hasil klasifikasi menunjukkan bahwa kernel *Polynomial* menghasilkan performa terbaik dengan nilai *accuracy*, *precision*, *recall*, dan *F1-score* sebesar 100%. Hasil penelitian menunjukkan bahwa integrasi *K-Means* dan *Multi-Class SVM* mampu mengidentifikasi tingkat kesadaran keamanan informasi secara efektif serta memberikan pendekatan yang lebih adaptif dibandingkan pengelompokan berdasarkan ambang skor yang telah ditentukan sebelumnya.

**Abstract:** The rapid development of information technology has increased data security risks, particularly in government institutions that manage strategic information. This study aims to apply machine learning methods to identify the information security awareness levels of Kendari City Government employees using the *Human Aspects of Information Security Questionnaire* (HAIS-Q). The *K-Means* algorithm was employed to form awareness categories based on data characteristics, while *Multi-Class Support Vector Machine* (SVM) was used to develop the classification model. Cluster evaluation using the *Elbow Method*, *Silhouette Score*, and *Davies-Bouldin Index* (DBI) identified three optimal clusters, with a *Silhouette Score* of 0.428 and a DBI value of 0.769. The classification results showed that the *Polynomial* kernel achieved the best performance, with 100% accuracy, precision, recall, and F1-score. The findings indicate that the integration of *K-Means* and *Multi-Class SVM* effectively identifies information security awareness levels and provides a more adaptive approach than categorization based on predefined score thresholds.

---

#### Corresponding Author:

Fiqram Faathir  
Email: [fiqramfaathir336@gmail.com](mailto:fiqramfaathir336@gmail.com)

---

## PENDAHULUAN

Indonesia merupakan salah satu negara dengan tingkat penggunaan internet yang tinggi. Pada tahun 2025, tercatat sekitar 212 juta penduduk Indonesia yang memiliki akses internet, mencerminkan tingkat penetrasi sebesar 74,6%. Selain itu, tercatat sebanyak 143 juta pengguna media sosial, yang mencakup 50,2% dari seluruh populasi (Riyanto, 2025). Perkembangan pesat teknologi informasi dan komunikasi telah memudahkan akses terhadap informasi dari berbagai tempat dan kapan saja. Kemajuan tersebut membawa pengaruh yang signifikan dalam mengubah tatanan dan wajah dunia (Rohmah, 2022). Perkembangan teknologi juga perlu diiringi dengan tingkat kesadaran keamanan informasi yang tinggi, terutama pada instansi pemerintahan agar keamanan data dan informasi dapat terjaga. Keamanan informasi merupakan langkah atau tindakan untuk melindungi aset informasi agar terhindar dari berbagai potensi ancaman (Nurul et al., 2022). Ancaman tersebut dapat timbul akibat kurangnya kesadaran dan pengetahuan pengguna, seperti membagikan data pribadi, menjelajah internet tanpa kewaspadaan, menggunakan kata sandi yang lemah, atau mengakses informasi melalui banyak perangkat sekaligus. Dalam konteks keamanan informasi, manusia dianggap sebagai titik paling rentan karena dapat dimanfaatkan oleh pelaku *social engineering* untuk memperoleh akses atau informasi yang bersifat rahasia (Ramadhan & Purwandari, 2023).

Kesadaran keamanan informasi perlu menjadi perhatian utama di instansi pemerintahan untuk menjaga keamanan data pemerintahan dan informasi pribadi pegawai. Berdasarkan data Databoks, pada kuartal III tahun 2022 Indonesia berada di peringkat ketiga sebagai negara dengan jumlah kasus kebocoran data terbanyak di dunia. Selama periode tersebut, sebanyak 12.742.031 akun di Indonesia dilaporkan terdampak insiden kebocoran data. Pada 20 Juni 2024, Kementerian Komunikasi dan Informatika (Kominfo) bersama Badan Siber dan Sandi Negara (BSSN) melaporkan terjadinya serangan siber yang menargetkan Pusat Data Nasional Sementara (PDNS). Insiden tersebut menunjukkan bahwa Indonesia masih menghadapi tantangan dalam memperkuat sistem keamanan siber guna melindungi data dan layanan digital nasional. Peristiwa tersebut menunjukkan pentingnya meningkatkan kesadaran keamanan informasi sebagai salah satu upaya dalam meminimalkan risiko kebocoran data pada instansi pemerintahan (Sukman, 2024).

Identifikasi tingkat kesadaran keamanan informasi pegawai pemerintah menjadi langkah penting untuk mengetahui tingkat pengetahuan, sikap, dan perilaku pegawai terhadap keamanan informasi. Berdasarkan hasil wawancara peneliti dengan pihak Dinas Komunikasi dan Informatika Kota Kendari, Pemerintah Kota Kendari masih menghadapi berbagai serangan siber sehingga identifikasi tingkat kesadaran keamanan informasi pegawai menjadi langkah awal dalam mendukung manajemen risiko keamanan informasi. *Human Aspects of Information Security-Questionnaire (HAIS-Q)* yang dikembangkan mengacu pada kerangka KAB yang digunakan untuk mengukur tiga aspek utama, yaitu pengetahuan (*Knowledge*), sikap (*Attitude*), dan perilaku (*Behaviour*) (Gofur et al., 2024). Dalam kerangka tersebut, *HAIS-Q* mencakup tujuh area utama, yaitu pengelolaan kata sandi, penggunaan *e-mail*, penggunaan internet, penggunaan jejaring sosial, pelaporan insiden, komputasi seluler, dan penanganan informasi. Setiap area utama tersebut terbagi menjadi tiga sub-area, sehingga keseluruhan model terdiri atas 21 sub-area yang merepresentasikan ketujuh area utama tersebut (Dewi et al., 2024).

Penelitian terdahulu telah banyak memanfaatkan *Human Aspects of Information Security Questionnaire (HAIS-Q)* untuk mengukur tingkat kesadaran keamanan informasi pada berbagai kelompok responden, seperti evaluasi efektivitas gamifikasi dalam meningkatkan kesadaran keamanan informasi (Wu et al., 2021), maupun pengukuran tingkat kesadaran keamanan informasi pada berbagai organisasi (Gofur et al., 2024; Kusumaningrum et al., 2022; Ramadhan & Purwandari, 2023). Pada umumnya, tingkat kesadaran keamanan informasi dengan pendekatan *HAIS-Q* ditentukan berdasarkan ambang skor yang telah dikelompokkan ke dalam kategori rendah, sedang, dan tinggi (Gofur et al., 2024; Kusumaningrum et al., 2022; Ramadhan & Purwandari, 2023). Meskipun pendekatan tersebut mampu memberikan gambaran tingkat kesadaran keamanan informasi, penentuan kategori berdasarkan ambang skor tersebut memiliki keterbatasan karena ambang skor tersebut belum mempertimbangkan karakteristik atau pola alami data responden berdasarkan 7 aspek keamanan informasi. Akibatnya, kategori yang dihasilkan berpotensi kurang merepresentasikan kondisi

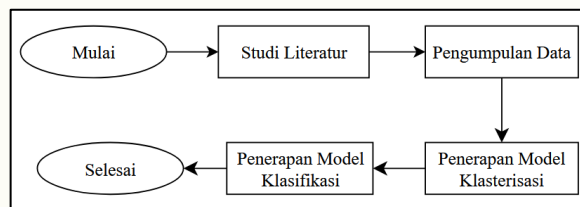
sebenarnya dari tingkat kesadaran keamanan informasi. Oleh karena itu, diperlukan pendekatan berbasis data yang mampu membentuk kategori berdasarkan karakteristik data responden.

Seiring perkembangan teknologi, *machine learning* telah banyak diterapkan untuk menyelesaikan berbagai permasalahan, seperti klasifikasi dan klusterisasi pada berbagai bidang (Bughin et al., 2025; Han, 2025; Sakalessy & Purnomo, 2024). Dalam bidang keamanan informasi, Saridewi & Sari (2021) menunjukkan bahwa algoritma *machine learning* dapat diterapkan untuk melakukan klusterisasi dan klasifikasi tingkat kesadaran keamanan informasi. Namun, penelitian tersebut menggunakan setiap butir pertanyaan *HAIS-Q* sebagai fitur masukan sehingga representasi data masih berada pada tingkat butir pertanyaan (item-level), bukan pada tujuh area utama *HAIS-Q*. Berdasarkan pendekatan tersebut, penelitian ini terlebih dahulu membentuk tujuh fitur berdasarkan tujuh area utama *HAIS-Q* sebelum dilakukan proses klusterisasi menggunakan *K-Means* dan klasifikasi menggunakan *Multi-Class Support Vector Machine* (SVM). Pendekatan ini memungkinkan proses identifikasi dilakukan pada tingkat aspek kesadaran keamanan informasi sehingga hasil analisis dapat diinterpretasikan secara langsung pada setiap area kesadaran keamanan informasi. Penelitian ini diharapkan dapat memberikan pendekatan alternatif dalam identifikasi tingkat kesadaran keamanan informasi pegawai pemerintah melalui integrasi *K-Means* dan *Multi-Class SVM* dengan representasi tujuh area *HAIS-Q* untuk membentuk kategori berdasarkan karakteristik data, sehingga proses identifikasi menjadi lebih adaptif dan mudah diinterpretasikan dibandingkan pengelompokan berdasarkan ambang skor yang telah ditentukan.

## METODE

### Alur Penelitian Secara Umum

Alur penelitian secara umum merupakan tahapan penelitian secara terstruktur mulai dari awal hingga akhir. Berikut adalah alur penelitian secara umum yang dapat dilihat pada Gambar 1:

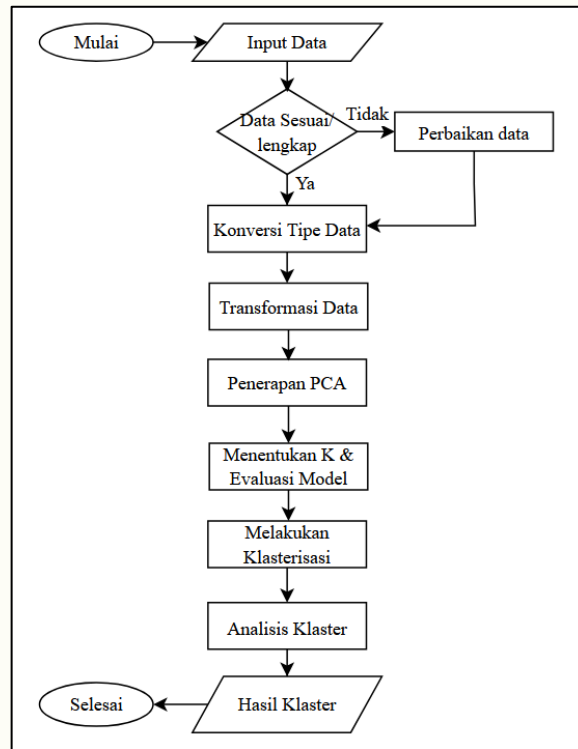


Gambar 1. Alur Penelitian Secara Umum

Berikut merupakan penjelasan dari alur penelitian secara umum yang ditampilkan pada Gambar 1: **Mulai:** Titik awal dari proses penelitian, **Studi Literatur:** Pada tahapan studi literatur akan dilakukan eksplorasi untuk mendapatkan pengetahuan dan pemahaman yang mendalam tentang kesadaran keamanan informasi pegawai, pendekatan yang digunakan, serta penerapan *machine learning* dalam klusterisasi dan klasifikasi tingkat kesadaran keamanan informasi, **Pengumpulan data:** Pengumpulan data dilakukan melalui penyebaran kuesioner yang mengadopsi pendekatan *Human Aspects of Information Security Questionnaire (HAIS-Q)* untuk memperoleh data mengenai tingkat kesadaran keamanan informasi pegawai. Responden dalam penelitian ini adalah pegawai pemerintah yang bertugas di Balai Kota Kendari, **Penerapan model Klusterisasi:** Pada tahap ini dilakukan reduksi dimensi menggunakan *Principal Component Analysis (PCA)*, kemudian diterapkan algoritma *K-Means* untuk membentuk kategori tingkat kesadaran keamanan informasi. Jumlah kluster optimal ditentukan berdasarkan evaluasi *Elbow Method*, *Silhouette Score*, dan *Davies-Bouldin Index (DBI)*. Selanjutnya dilakukan identifikasi karakteristik setiap kluster berdasarkan nilai *centroid* pada masing-masing aspek *HAIS-Q*, **Penerapan Model Klasifikasi:** Hasil klusterisasi digunakan sebagai label dalam proses klasifikasi menggunakan algoritma *Multi-Class (SVM)*. Selanjutnya, model dievaluasi menggunakan *confusion matrix* berdasarkan metrik *accuracy*, *precision*, *recall*, dan *F1-score*, serta diuji menggunakan data validasi untuk mengetahui kemampuan model dalam mengklasifikasikan data baru, dan **Selesai:** Titik akhir dari proses.

### Alur Model Klasterisasi

Berikut dibawah ini adalah alur model klasterisasi tingkat kesadaran kemananan informasi dengan menggunakan algoritma *K-Means* yang dapat dilihat pada Gambar 2:

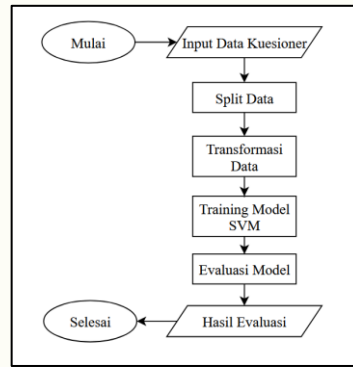


Gambar 2. Alur Pengembangan Model Klasterisasi

Berikut merupakan penjelasan dari alur pengembangan model klasterisasi yang ditampilkan pada Gambar 2: Mulai: Titik awal dari proses, *Input Data*: Data yang akan di klaster diinputkan pada model, setelah itu diperikasa apakah seluruh data lengkap/sesuai. Jika terdapat data yang hilang atau tidak sesuai dengan *input* yang diperlukan seperti simbol dan lainnya akan di perbaiki sebelum masuk ke tahapan selanjutnya, *Konversi Tipe Data*: Pada tahapan ini akan mengubah data fitur dari *object* ke *float*, *Transformasi Data*: Nilai yang diperoleh dari hasil kuesioner kemudian ditransformasikan dari skala 1–5 ke dalam rentang 0–100. Proses ini bertujuan untuk menyeragamkan skala data sebelum dilakukan analisis lebih lanjut, *Penerapan PCA (Principal Component Analysis)*: Tahapan ini dilakukan untuk mereduksi dimensi fitur sebelum dilakukan tahapan selanjutya, *Menentukan nilai K & evaluasi Klaster*: Tahapan ini bertujuan untuk menentukan jumlah klaster (K) pada algoritma *K-Means* berdasarkan 3 metode evaluasi klaster yaitu *Elbow Method*, *Silhouette Score*, dan *DBI*. Nilai K yang optimal berdasarkan 3 metode evaluasi akan digunakan dalam tahapan selanjutnya untuk analisis klister, *Melakukan Klasterisasi*: Tahapan klasterisasi menggunakan algoritma *K-Means* dilakukan dengan menggunakan nilai K yang optimal berdasarkan evaluasi klaster sehingga jumlah klaster yang terbentuk sebanyak K klister, *Analisis klaster*: Hasil klaster kemudian dianalisis untuk mengetahui karakteristik dari tiap klaster dan output tingkat kesadaran keamanan informasi pada tiap klister, *Hasil klaster*: *Output* yang dihasilkan pada tahapan ini berupa hasil klasterisasi tingkat kesadaran keamanan informasi pegawai pemerintah, dan *Selesai*: Titik akhis dari proses.

### Alur Model Klasifikasi

Diagram alur pembuatan model klasifikasi tingkat kesadaran keamanan informasi dengan menggunakan algoritma *Multi-Class SVM* dapat dilihat pada Gambar 3:



Gambar 3. Alur Model Klasifikasi

Berikut merupakan penjelasan dari alur pembuatan model klasifikasi yang ditampilkan pada Gambar 3: *Mulai*: Titik awal dari proses, *Input data kuesioner*: *Dataset* kuesioner yang diinputkan adalah data kuesioner tingkat kesadaran keamanan informasi pegawai pemerintah Kota Kendari yang telah dikumpulkan dan telah diberi label berdasarkan hasil kluster, *Split data*: *Split* data dilakukan dengan membagi *dataset* menjadi data *training* dan data *testing*, *Transformasi data*: Transform data untuk mengonversi data dari rentang 1-5 menjadi rentang 0-100, *Training Model SVM*: *Training model* dengan menggunakan algoritma *Multi-Class SVM* dengan membandingkan antara 4 kernel, yaitu *Linear*, *Polynomial*, *Sigmoid*, dan *RBF*, *Hasil Evaluasi*: Hasil evaluasi yang didapatkan berupa *accuracy*, *precision*, *recall*, dan *F1-score* dari tiap kernel, dan *Selesai*: Titik akhir dari proses.

#### **Human Aspects of Information Security Questionnaire (HAIS-Q)**

Pengukuran tingkat kesadaran individu terhadap keamanan informasi dalam penelitian ini menggunakan *Human Aspects of Information Security Questionnaire (HAIS-Q)*, yaitu instrumen yang telah divalidasi pada penelitian sebelumnya. Instrumen ini mencakup tujuh aspek utama, yaitu pengelolaan kata sandi, penggunaan *e-mail*, akses dan pemanfaatan internet, aktivitas di media sosial, penggunaan perangkat seluler, pengendalian informasi, serta prosedur pelaporan insiden. Seluruh aspek tersebut selanjutnya diklasifikasikan ke dalam tiga dimensi utama, yakni *Knowledge*, *Attitude*, dan *Behaviour* (*KAB*) (Fadlika et al., 2023). Kuesioner tingkat kesadaran keamanan informasi pegawai pemerintah dengan pendekatan *HAIS-Q* pada penelitian ini mewakili tujuh aspek *HAIS-Q* dan model *Knowledge*, *Attitude*, *Behaviour* (*KAB*) dengan total 20 pertanyaan yang disajikan pada Tabel 2:

Tabel 1. Pertanyaan Kuesioner *HAIS-Q*

| No | Pertanyaan                                                                                                                                                    | Sub-Area                     | Skala Liekrt        |
|----|---------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------|---------------------|
| 1. | Saya mengetahui bahwa kata sandi harus menggunakan kata sandi yang unik dan kuat untuk perangkat kerja (laptop/komputer/akun aplikasi/dll) (pengetahuan)      | Mengunci <i>workstation</i>  | Tingkat Pengetahuan |
| 2. | Saya merasa aman membagikan kata sandi perangkat ataupun akun saya terutama melalui chat/telepon (sikap)                                                      | Berbagi kata sandi           | Tingkat Persetujuan |
| 3. | Saya menggunakan karakter yang unik (huruf, angka, simbol) dalam membuat kata sandi (perilaku)                                                                | Memilih kata sandi yang baik | Frekuensi           |
| 4. | Saya mengetahui bahwa membagikan tautan dalam <i>e-mail</i> /SMS dari pengirim terkenal atau mencurigakan berpotensi menyebarkan serangan virus (pengetahuan) | Meneruskan <i>e-mail</i>     | Tingkat Pegetahuan  |
| 5. | Saya merasa aman jika saya mengakses tautan dalam <i>e-mail</i> /SMS dari orang yang tidak dikenal (sikap)                                                    | Membuka lampiran             | Tingkat Persetujuan |

|     |                                                                                                                                                                        |                                                       |                     |
|-----|------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------|---------------------|
| 6.  | Saya berinisiatif untuk melaporkan kepada departemen IT/pihak IT di lingkungan kerja saya jika mendapatkan e-mail yang mencurigakan (perilaku)                         | Tanggung jawab departemen IT                          | Tingkat Persetujuan |
| 7.  | Saya mengetahui bahwa berbahaya mengakses situs dewasa, situs perjudian, situs unduh berbasis <i>torrent</i> , atau situs sejenis lainnya (pengetahuan)                | Penggunaan internet yang tidak pantas                 | Tingkat pengetahuan |
| 8.  | Saya merasa penting memastikan keamanan suatu situs web ( <i>URL, HTTPS, VPN</i> , sertifikat digital, dan lainnya) sebelum memberikan informasi secara daring (sikap) | Mengakses situs yang meragukan                        | Tingkat Persetujuan |
| 9.  | Saya mengunduh file/aplikasi di perangkat dari situs web resmi/terpercaya dan menghindari mengunduh dari situs ilegal/bajakan (perilaku)                               | Menginstal perangkat lunak tanpa izin                 | Frekuensi           |
| 10. | Saya mengetahui bahwa penggunaan sosial media yang berlebihan di waktu kerja berpotensi memberi celah pada serangan siber (pengetahuan)                                | Waktu kerja yang dihabiskan untuk SNS                 | Tingkat Pengetahuan |
| 11. | Saya merasa unggahan/postingan di media sosial tidak akan memberikan dampak apa pun terhadap keamanan informasi                                                        | Penggunaan Situs Jejaring Sosial                      | Tingkat Persetujuan |
| 12. | Saya menghindari mengunggah/ memposting informasi pribadi (KTP, pelat nomor kendaraan, dan lainnya) atau data terkait pekerjaan di media sosial (perilaku)             | Membuat postingan tentang pekerjaan di Media Sosial   | Frekuensi           |
| 13. | Saya akan melaporkan kepada pihak berwenang tindakan mencurigakan terkait keamanan informasi yang saya lihat meskipun dilakukan oleh orang yang dikenal (sikap).       | Melaporkan individu yang mencurigakan                 | Tingkat Persetujuan |
| 14. | Saya pernah membuat laporan insiden keamanan kepada pihak berwenang seperti di media sosial, pengaduan ke BRTI, Kepolisian, OJK, lapor.go.id,                          | Melaporkan semua insiden keamanan                     | Frekuensi           |
| 15. | Saya mengetahui bahwa mengirimkan informasi melalui jaringan seluler beresiko membuka celah untuk mencuri informasi (pengetahuan)                                      | Mengirimkan informasi sensitif                        | Tingkat Pengetahuan |
| 16. | Saya merasa khawatir ketika meninggalkan perangkat kerja/laptop saya di meja kerja dalam keadaan terbuka tanpa diawasi (sikap)                                         | Mengamankan secara fisik perangkat elektronik pribadi | Tingkat Persetujuan |
| 17. | Saya pernah menggunakan Wi-Fi gratis di tempat umum untuk mengakses data penting (transaksi keuangan, <i>e-mail</i> pekerjaan, dan lainnya) (perilaku)                 | Memeriksa e-mail kantor melalui jaringan gratis       | Frekuensi           |
| 18. | Saya mengetahui bahwa <i>USB/drive</i> dari orang yang tak dikenal atau tidak diketahui asalnya beresiko menyebarkan virus (pengetahuan)                               | Menyisipkan perangkat DVD/USB                         | Tingkat Pengetahuan |
| 19. | Saya merasa aman ketika membuang file/dokumen kerja langsung ke tempat sampah atau <i>recycle bin</i> (sikap)                                                          | Membuang dokumen sensitif                             | Tingkat Persetujuan |
| 20. | Saya selalu menutup dokumen kerja yang sensitif/penting setelah dibuka atau digunakan (perilaku)                                                                       | Meninggalkan materi sensitif yang telah digunakan     | Frekuensi           |

### Pengumpulan Sampel dan *Margin of Error*

Sampel merupakan sebagian dari populasi yang digunakan sebagai sumber data utama dalam suatu penelitian. Jumlah sampel ditentukan menggunakan rumus *Slovin* sebagaimana ditunjukkan pada Persamaan (1) (Amin et al., 2023).

$$n = \frac{N}{1 + N(e)^2} \quad (1)$$

Keterangan:  $n$  = Banyaknya Sampel

$N$  = Banyaknya Populasi

$e$  = *Margin of Error*

Penelitian ini menggunakan teknik pengambilan sampel secara acak dengan jumlah 243 responden. Populasi penelitian terdiri atas 894 Pegawai Pemerintah Kota Kendari yang diperoleh dari PNS (termasuk CPNS) dan PPPK (tidak termasuk PPPK Paruh Waktu) pada seluruh instansi di Balai Kota Kendari. Berdasarkan rumus *Slovin*, jumlah sampel tersebut menghasilkan *margin of error* 5,5%.

### Uji Validitas dan Reliabilitas

Uji validitas dan reliabilitas dilakukan untuk memastikan bahwa instrumen penelitian layak digunakan dalam pengumpulan data. Uji validitas bertujuan mengevaluasi kemampuan setiap butir pertanyaan dalam mengukur variabel yang diteliti, di mana suatu item dinyatakan valid apabila nilai  $r$  hitung melebihi  $r$  tabel. Selanjutnya, uji reliabilitas dilakukan menggunakan metode *Cronbach's Alpha* untuk menilai tingkat konsistensi instrumen. Instrumen dinyatakan reliabel jika memiliki nilai *Cronbach's Alpha* sebesar 0,70 atau lebih, yang menunjukkan bahwa instrumen mampu menghasilkan pengukuran yang konsisten (Meivira et al., 2022).

### Principal Component Analysis

*Principal Component Analysis* merupakan metode yang digunakan untuk mereduksi dimensi data dengan mengubah sejumlah variabel menjadi beberapa komponen utama yang tetap mempertahankan sebagian besar informasi pada data (Prasetya et al., 2026). Pada penelitian ini, *PCA* digunakan sebagai tahap praproses untuk mereduksi tujuh fitur *HAIS-Q* menjadi dua komponen utama sebelum dilakukan proses klusterisasi menggunakan *K-Means*. Penggunaan *PCA* bertujuan untuk mengurangi kompleksitas data sehingga proses klusterisasi menjadi lebih efisien dengan tetap mempertahankan sebagian besar informasi pada data.

### Transformasi Data

Transformasi data dilakukan dengan mengonversi skala dari nilai tiap aspek/fitur pada kuesioner dari rentang skala *likert* 1 sampai 5 menjadi rentang nilai 0 – 100 dengan menggunakan Persamaan (9).

$$\text{Skala 100} = \frac{\text{Skor Likert} - 1}{5 - 1} \times 100 \quad (2)$$

Ket: *Skor Likert* = nilai fitur/aspek ( $a$ )

### *K-Means Clustering*

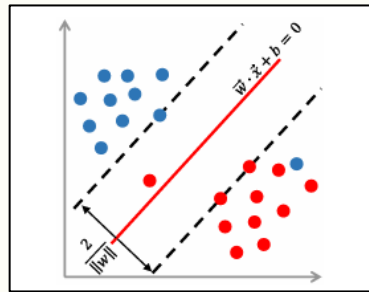
Sebagai algoritma *unsupervised learning*, *K-Means* membentuk kelompok data berdasarkan kesamaan karakteristik antar data. Setiap kelompok direpresentasikan oleh sebuah pusat kluster (*centroid*), sementara proses optimasinya dilakukan dengan meminimalkan jumlah kuadrat jarak antara seluruh data dan *centroid* dari kluster tempat data tersebut berada (Khan et al., 2021). Pada penelitian ini, algoritma *K-Means* diterapkan pada data hasil reduksi dimensi menggunakan *PCA* untuk mengelompokkan respon kuesioner berdasarkan pola karakteristik jawaban responden. Hasil pengelompokan tersebut kemudian dianalisis untuk memperoleh kategori tingkat kesadaran keamanan informasi berdasarkan karakteristik aspek *HAIS-Q* masing-masing kluster. Jumlah kluster ditentukan melalui evaluasi beberapa nilai  $k$  menggunakan *Elbow Method*, *Silhouette Score* dan *Davies-Bouldin Index*. Selanjutnya, hasil klusterisasi digunakan sebagai label kelas dalam proses klasifikasi menggunakan algoritma *Multi-Class SVM*.

### Split Data

*Dataset* hasil klusterisasi yang telah melalui tahap transformasi selanjutnya dipisahkan menjadi data *training* dan *testing* dengan rasio 70:30. Proses pelatihan model *Multi-Class SVM* dilakukan menggunakan data *training*, sedangkan pengujian kinerja model memanfaatkan data *testing* melalui *confusion matrix* berdasarkan metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Sebagai evaluasi tambahan, penelitian ini menggunakan 23 data validasi di luar data pelatihan dan pengujian untuk menilai kemampuan model dalam mengklasifikasikan data baru.

### Multi-Class Support Vector Machine (SVM)

*Support Vector Machine (SVM)* merupakan algoritma *supervised learning* yang banyak digunakan untuk menyelesaikan permasalahan klasifikasi. Algoritma ini menerapkan prinsip *Structural Risk Minimization (SRM)* untuk menghasilkan model dengan kemampuan generalisasi yang baik melalui pengendalian kompleksitas model selama proses pembelajaran. (Manik et al., 2021). SVM mengklasifikasikan data dengan mencari *hyperplane* yang menghasilkan margin maksimum antar kelas. *Hyperplane* merupakan batas pemisah yang digunakan untuk membedakan atau memisahkan data ke dalam kelas berdasarkan label. *Hyperplane* yang optimal diperoleh dengan menghitung margin antara kelas-kelas tersebut dan memilih nilai margin yang paling besar (Taufiqurrahman et al., 2021). Vektor pendukung (*support vector*) adalah titik data yang paling dekat dengan *hyperplane* dan membantu menentukan di mana letaknya. Disebut *support vector* karena titik-titik tersebut berperan dalam menentukan posisi *hyperplane* optimal. Ilustrasi sederhana SVM dapat dilihat pada Gambar 4.



Gambar 4 Ilustrasi SVM (Liu et al., 2019)

SVM digunakan untuk membedakan data ke dalam dua kategori dengan tujuan menemukan *hyperplane* yang memiliki margin paling optimal. Namun, ketika jumlah kelas melebihi dua, diperlukan pendekatan khusus, salah satunya dengan menerapkan metode SVM multikelas seperti *One-versus-One (OvO)* dan *One-versus-All (OvA)* (Nurkholis et al., 2022). Pada penelitian ini, algoritma *Multi-Class SVM* digunakan untuk membangun model klasifikasi berdasarkan label kelas yang diperoleh dari hasil klusterisasi *K-Means*. Penggunaan label hasil klusterisasi bertujuan membentuk kategori tingkat kesadaran keamanan informasi berdasarkan karakteristik alami data responden, sehingga tidak bergantung pada pengelompokan menggunakan ambang skor yang telah ditentukan sebelumnya. Selanjutnya, model *Multi-Class SVM* dilatih untuk mempelajari pola dari hasil klusterisasi tersebut sehingga mampu mengklasifikasikan data kuesioner baru ke dalam kategori yang sama tanpa perlu melakukan proses klusterisasi ulang. Dengan demikian, model yang dikembangkan dapat mendukung proses identifikasi tingkat kesadaran keamanan informasi pegawai Pemerintah Kota Kendari secara lebih efisien dan konsisten pada pengumpulan data survei berikutnya.

### Evaluasi Model

Evaluasi model klusterisasi dilakukan menggunakan *Elbow Method*, *Silhouette Score*, dan *Davies-Bouldin Index (DBI)* untuk menentukan jumlah kluster optimal. Selanjutnya, performa model *Multi-Class SVM* pada setiap kernel dievaluasi menggunakan *confusion matrix* berdasarkan metrik *accuracy*, *precision*, *recall*, dan *F1-score*.

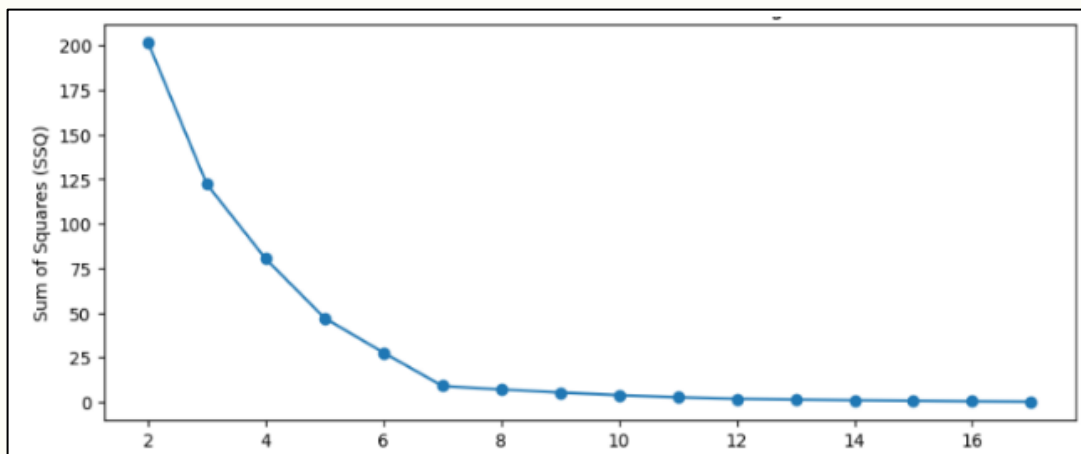
#### 1. Silhouette Score

*Silhouette Score* merupakan metrik yang digunakan dalam menilai kualitas hasil klusterisasi suatu data. Metrik ini mengukur tingkat kesesuaian suatu titik data terhadap kluster tempatnya berada serta

keterpisahannya dari kluster lain. Skor ini sering dimanfaatkan untuk mengevaluasi kinerja algoritma klusterisasi, seperti *K-Means*.

## 2. Elbow Method

*Elbow Method* adalah teknik penentuan jumlah kluster optimal yang dilakukan dengan mengamati penurunan nilai *Sum of Squared Error* (*SSE*) pada setiap jumlah kluster yang diuji. Inti dari metode ini adalah mengidentifikasi titik "elbow" atau siku pada grafik, yaitu titik ketika penurunan *SSE* mulai mengalami perlambatan signifikan. Titik tersebut menunjukkan bahwa penurunan nilai *SSE* mulai berkurang sehingga penambahan jumlah kluster tidak lagi memberikan peningkatan yang signifikan. Karena itu, titik siku tersebut dianggap sebagai jumlah kluster yang tepat (Sofia, 2023). Berikut adalah contoh grafik dari *Sum of Square Error* yang akan digunakan untuk melihat titik optimal kluster pada Gambar 4:



Gambar 5 Contoh Grafik *Elbow Method* (Sofia, 2023)

## 3. Davies-Bouldin Index

*Davies-Bouldin Index* (*DBI*) digunakan untuk mengevaluasi kualitas hasil klusterisasi dengan membandingkan nilai *DBI* pada beberapa jumlah kluster (*K*). Metode ini menilai kualitas kluster berdasarkan kemampuan meminimalkan jarak antar data dalam kluster yang sama serta memaksimalkan jarak antar kluster yang berbeda. Semakin kecil nilai *DBI* yang diperoleh, semakin baik kualitas klusterisasi yang dihasilkan (Az-Zahra et al., 2021).

## 4. Confusion Matrix

Salah satu metode yang umum digunakan untuk menilai performa model *machine learning* adalah *confusion matrix*. Metode ini juga berfungsi untuk mengidentifikasi permasalahan pada proses klasifikasi, khususnya ketika model menghasilkan output dengan dua atau lebih kelas (Monasari & Pratiwi, 2025). Evaluasi performa model *Multi-Class SVM* dilakukan menggunakan *confusion matrix* dengan empat metrik, yaitu *accuracy*, *precision*, *recall*, dan *F1-score*. *Accuracy* merepresentasikan proporsi prediksi yang benar, *precision* menunjukkan ketepatan hasil klasifikasi pada setiap klasifikasi positif, *recall* mengukur kemampuan model dalam mendeteksi data yang termasuk ke dalam kelas sebenarnya, sedangkan *F1-score* diukur melalui rata-rata harmonik *precision* dan *recall* (Islam et al., 2022).

# HASIL DAN PEMBAHASAN

## Uji Validitas dan Relabilitas

Sebelum dilakukan proses klusterisasi dan klasifikasi, instrumen penelitian terlebih dahulu diuji validitas dan reliabilitas untuk memastikan kelayakan instrumen yang digunakan.

### 1. Uji Validitas

Tabel uji validitas instrumen kuesioner dari pertanyaan 1 sampai pertanyaan 20 disajikan pada Tabel 2.

Tabel 2. Hasil Uji Validitas

| Pertanyaan (Q)       | Varian Butir  | r hitung | r tabel | Validitas |
|----------------------|---------------|----------|---------|-----------|
| Q1                   | 0,912         | 0,682    | 0,126   | Valid     |
| Q2                   | 1,142         | 0,478    | 0,126   | Valid     |
| Q3                   | 1,288         | 0,613    | 0,126   | Valid     |
| Q4                   | 1,29          | 0,645    | 0,126   | Valid     |
| Q5                   | 1,07          | 0,571    | 0,126   | Valid     |
| Q6                   | 0,868         | 0,578    | 0,126   | Valid     |
| Q7                   | 0,98          | 0,659    | 0,126   | Valid     |
| Q8                   | 0,661         | 0,700    | 0,126   | Valid     |
| Q9                   | 1,599         | 0,680    | 0,126   | Valid     |
| Q10                  | 1,590         | 0,640    | 0,126   | Valid     |
| Q11                  | 1,195         | 0,550    | 0,126   | Valid     |
| Q12                  | 2,719         | 0,465    | 0,126   | Valid     |
| Q13                  | 0,832         | 0,608    | 0,126   | Valid     |
| Q14                  | 0,904         | 0,293    | 0,126   | Valid     |
| Q15                  | 1,241         | 0,672    | 0,126   | Valid     |
| Q16                  | 1,064         | 0,611    | 0,126   | Valid     |
| Q17                  | 1,638         | 0,367    | 0,126   | Valid     |
| Q18                  | 1,235         | 0,695    | 0,126   | Valid     |
| Q19                  | 1,403         | 0,406    | 0,126   | Valid     |
| Q20                  | 1,367         | 0,651    | 0,126   | Valid     |
| <b>Varians Butir</b> | <b>24,997</b> |          |         |           |

Berdasarkan Tabel 3, seluruh butir pertanyaan memiliki nilai r hitung lebih besar daripada r tabel (0,126), sehingga seluruh 20 butir pertanyaan dinyatakan valid dan layak digunakan pada tahap analisis selanjutnya serta mempunyai varians butir 24,997.

## 2. Uji Relabilitas

Setelah seluruh butir pertanyaan dinyatakan valid, dilakukan uji reliabilitas menggunakan metode *Cronbach's Alpha* untuk mengukur tingkat konsistensi instrumen penelitian. Hasil uji reliabilitas disajikan pada Tabel 3.

| Tabel 3. Variabel Uji Relabilitas |         |
|-----------------------------------|---------|
| Variabel                          | Nilai   |
| Banyaknya pertanyaan              | 20      |
| Jumlah varians butir              | 24,997  |
| Varians total                     | 160,554 |
| <i>Cronbach's Alpha</i>           | 0,889   |

Berdasarkan Tabel 3, diperoleh nilai *Cronbach's Alpha* sebesar 0,889 yang menunjukkan bahwa instrumen penelitian memiliki tingkat reliabilitas yang sangat baik. Dengan demikian, instrumen layak digunakan pada proses analisis selanjutnya.

## Implementasi Model Klasterisasi

### 1. Dataset

*Dataset* yang digunakan pada proses klasterisasi merupakan data hasil kuesioner yang terdiri atas 243 responden, yaitu 192 PNS (termasuk CPNS) dan 51 PPPK.

### 2. Fitur Klasterisasi

Proses klasterisasi menggunakan tujuh fitur yang diperoleh dari nilai rata-rata setiap aspek *HAIS-Q*. Nilai setiap aspek dihitung dari rata-rata dua atau tiga butir pertanyaan yang merepresentasikan aspek tersebut. Pembentukan setiap fitur disajikan pada Tabel 3.

Tabel 4 Pembentuk Nilai Aspek (Fitur)

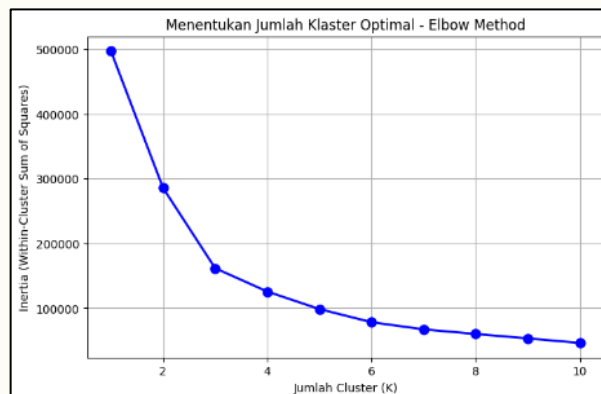
| No  | Pertanyaan (Q)                                                     | Aspek/Fitur (A)                                      |
|-----|--------------------------------------------------------------------|------------------------------------------------------|
| 1.  | Q1/A1.1 (Mengunci <i>workstation</i> )                             | A1 (Pengelolaan Kata Sandi)                          |
| 2.  | Q2/A1.2 (Berbagi kata sandi)                                       |                                                      |
| 3.  | Q3/A1.3 (Memilih kata sandi yang baik)                             |                                                      |
| 4.  | Q4/A2.1 (Meneruskan <i>e-mail</i> )                                | A2 (Penggunaan <i>E-mail</i> )                       |
| 5.  | Q5/A2.2 (Membuka lampiran)                                         |                                                      |
| 6.  | Q6/A2.3 (Tanggung jawab departemen IT)                             |                                                      |
| 7.  | Q7/A3.1 (Penggunaan internet yang tidak pantas)                    | A3 (Penggunaan Internet)                             |
| 8.  | Q8/A3.2 (Mengakses situs yang meragukan)                           |                                                      |
| 9.  | Q9/A3.3 (Menginstal perangkat lunak tanpa izin)                    |                                                      |
| 10. | Q10/A4.1 (Waktu kerja yang dihabiskan untuk <i>SNS</i> )           | A4 (Penggunaan Situs Jejaring Sosial ( <i>SNS</i> )) |
| 11. | Q11/A4.2 (Penggunaan Situs Jejaring Sosial ( <i>SNS</i> ))         |                                                      |
| 12. | Q12/A4.3 (Membuat postingan tentang pekerjaan di <i>SNS</i> )      |                                                      |
| 13. | Q13/A5.1 (Melaporkan individu yang mencurigakan)                   | A5 (Pelaporan Insiden)                               |
| 14. | Q14/A5.2 (Melaporkan semua insiden keamanan)                       |                                                      |
| 15. | Q15/A6.1 (Mengirimkan informasi sensitif melalui jaringan seluler) | A6 (Komputasi Seluler)                               |
| 16. | Q16/A6.2 (Mengamankan secara fisik perangkat elektronik pribadi)   |                                                      |
| 17. | Q17/A6.3 (Memeriksa e-mail kantor melalui jaringan gratis)         |                                                      |
| 18. | Q18/A7.1 (Menyisipkan perangkat <i>DVD/USB</i> )                   | A7 (Penanganan Informasi)                            |
| 19. | Q19/A7.2 (Membuang dokumen sensitif)                               |                                                      |
| 20. | Q20/A7.3 (Meninggalkan materi sensitif yang telah digunakan)       |                                                      |

### 3. Principal Component Analysis (PCA)

*Principal Component Analysis (PCA)* digunakan untuk mereduksi dimensi data dari tujuh fitur *HAIS-Q* menjadi dua komponen utama. Kedua komponen tersebut mampu merepresentasikan 68,34% variasi data sehingga digunakan sebagai masukan pada proses klusterisasi *K-Means*. Selanjutnya, hasil klusterisasi dievaluasi menggunakan *elbow method*, *Silhouette Score*, dan *Davies-Bouldin Index* untuk menentukan jumlah kluster terbaik.

### 4. Elbow Method

Proses penentuan kluster optimal dengan menggunakan *elbow method* dimulai dengan melihat nilai *Sum Square of Error (SSE)* tiap nilai *K* dari rentang *K*=2 hingga *K*=11 kemudian diamati titik siku (*elbow*) yang paling signifikan ada pada *K* keberapa yang ditampilkan pada Gambar 6.



Gambar 6 Grafik Hasil Elbow Method

Grafik *Elbow Method* menunjukkan kluster optimal yang terbentuk berdasarkan titik siku (*elbow*) adalah 3 kluster (K=3).

##### 5. *Silhouette Score & Davies Boulin Index (DBI)*

Nilai *Silhouette Score* dan *Davies Bouldin Index* pada tiap K pada proses kluster dengan algoritma *K-Means* dapat dilihat pada Tabel 4.

Tabel 5. Nilai *Silhouette Score* dan *DBI*

| No. | Nilai K | <i>Silhouette Score</i> | <i>Davies Bouldin Index (DBI)</i> |
|-----|---------|-------------------------|-----------------------------------|
| 1.  | 2       | 0.392                   | 0.916                             |
| 2.  | 3       | 0.428                   | 0.769                             |
| 3.  | 4       | 0.344                   | 0.871                             |
| 4.  | 5       | 0.373                   | 0.827                             |
| 5.  | 6       | 0.385                   | 0.761                             |
| 6.  | 7       | 0.362                   | 0.808                             |
| 7.  | 8       | 0.374                   | 0.795                             |
| 8.  | 9       | 0.374                   | 0.825                             |
| 9.  | 10      | 0.362                   | 0.757                             |
| 10  | 11      | 0.364                   | 0.749                             |

Berdasarkan hasil evaluasi kluster menggunakan tiga metode, *Elbow Method* dan *Silhouette Score* menunjukkan bahwa jumlah kluster yang optimal adalah K = 3, sedangkan *Davies-Bouldin Index (DBI)* menunjukkan nilai optimal pada K = 11. Pada penelitian ini dipilih K = 3 sebagai jumlah kluster yang digunakan karena didukung oleh dua dari tiga metode evaluasi, sehingga dianggap lebih representatif dalam menggambarkan karakteristik data.

##### 6. Rata-rata Fitur Tiap Kluster (*Centroid*)

Tingkat Kesadaran Keamanan Informasi dibagi atas 3 rentang yaitu tinggi, sedang, dan rendah yang dapat dilihat pada Tabel 5.

Tabel 6. Rentang Nilai Tingkat Kesadaran Keamanan Informasi

| Tingkatan | Rentang     |
|-----------|-------------|
| Tinggi    | 80-100      |
| Sedang    | 60>=skor<80 |
| Rendah    | T<60        |

Untuk memudahkan interpretasi nilai *centroid* pada setiap aspek, digunakan acuan kategori skor, yaitu kategori tinggi berada pada rentang 80–100, kategori sedang pada rentang 60 hingga di bawah 80, dan kategori rendah pada rentang 0 hingga di bawah 60. Rata-rata nilai setiap fitur (*centroid*) pada masing-masing kluster beserta hasil klusterisasi disajikan pada Tabel 6.

Tabel 7. Rata-rata Tiap Fitur (*Centroid*) & Hasil Kluster

| No | Kluster | A1    | A2    | A3    | A4    | A5    | A6    | A7    |
|----|---------|-------|-------|-------|-------|-------|-------|-------|
| 1. | 1       | 73,11 | 68,29 | 69,14 | 44,86 | 37,79 | 59,57 | 56,84 |
| 2. | 2       | 25,76 | 18,94 | 16,67 | 13,64 | 12,5  | 19,7  | 9,09  |
| 3. | 3       | 88,94 | 83,81 | 89,5  | 73,56 | 49,52 | 77,72 | 75,72 |

Berdasarkan Tabel 7, hasil klusterisasi tidak hanya menghasilkan kategori tingkat kesadaran keamanan informasi responden, tetapi juga memberikan gambaran karakteristik setiap kluster berdasarkan nilai *centroid* pada tujuh aspek *HAIS-Q*. Dengan demikian, penelitian ini tidak hanya mengidentifikasi kategori tingkat kesadaran keamanan informasi, tetapi juga mampu menunjukkan aspek-aspek *HAIS-Q* yang menjadi kekuatan maupun kelemahan pada setiap kelompok responden.

## 7. Karakteristik dan Rekomendari Tiap Klaster (Label)

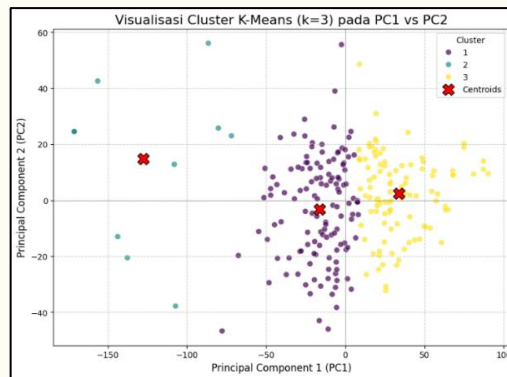
Label setiap klaster ditentukan berdasarkan karakteristik nilai *centroid* pada masing-masing aspek *HAIS-Q*. Keterangan setiap klaster beserta rekomendasi yang diusulkan disajikan pada Tabel 8.

Tabel 8. Keterangan Tiap Klaster (Label)

| No | Klaster | Keterangan                                                                                                                                                                                                                                                                  |
|----|---------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1. | 1       | Kesadaran keamanan informasi berada pada kategori sedang dengan kelemahan pada aspek penggunaan situs jejaring sosial, pelaporan insiden, komputasi seluler, dan penanganan informasi. Direkomendasikan pelatihan atau sosialisasi yang berfokus pada aspek-aspek tersebut. |
| 2. | 2       | Kesadaran keamanan informasi berada pada kategori rendah pada seluruh aspek. Direkomendasikan program pelatihan dasar keamanan informasi secara menyeluruh untuk meningkatkan pengetahuan, sikap, dan perilaku pegawai.                                                     |
| 3. | 3       | Kesadaran keamanan informasi berada pada kategori tinggi dengan kelemahan relatif pada aspek pelaporan insiden. Direkomendasikan edukasi dan penguatan prosedur pelaporan insiden keamanan informasi.                                                                       |

## 8. Plot Sebaran Klaster

Plot sebaran 3 klaster dengan menggunakan *Principal Component Analysis (PCA)* 2 fitur disajikan pada Gambar 7:

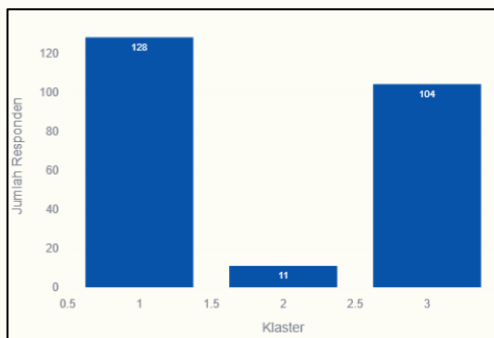


Gambar 7. Plot Sebaran Klaster

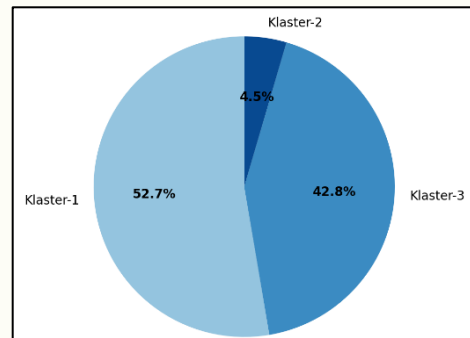
Visualisasi Plot sebaran menunjukkan distribusi klaster yang lebih menyebar pada klaster-2 sedangkan pada klaster-1 dan klaster-3 lebih terkonsentrasi dan terdapat beberapa titik data yang saling bertumpang tindih. Secara keseluruhan, visualisasi tersebut menunjukkan bahwa antar klaster memiliki pemisah yang cukup jelas.

## 9. Presentase klaster

Distribusi jumlah responden pada setiap klaster disajikan pada diagram batang Gambar 8. Sementara itu, persentase jumlah responden pada masing-masing klaster ditunjukkan diagram lingkaran pada Gambar 9.



Gambar 8. Distribusi Responden Tiap Klaster



Gambar 9. Persentase Responden Tiap Klaster



Berdasarkan hasil evaluasi, algoritma *Multi-Class SVM* menggunakan empat kernel, kernel *Polynomial* menghasilkan performa terbaik dengan akurasi sebesar 100%, *precision* 100%, *recall* 100%, dan *F1-score* 100%. Hasil tersebut menunjukkan bahwa kernel *Polynomial* lebih sesuai dengan karakteristik data pada penelitian ini sehingga menghasilkan performa klasifikasi yang lebih baik dibandingkan kernel lainnya. Kernel *Linear* memperoleh akurasi sebesar 98,48%, diikuti oleh kernel *RBF* sebesar 96,97%. Sementara itu, kernel *Sigmoid* menghasilkan performa terendah dengan akurasi sebesar 56,06%. Perbedaan performa antar kernel menunjukkan bahwa pemilihan fungsi kernel berpengaruh terhadap kemampuan *SVM* dalam membentuk *hyperplane* pemisah antar kelas. Model *Multi-Class SVM* mampu mempelajari pola hasil klusterisasi *K-Means* sehingga dapat mengklasifikasikan kategori tingkat kesadaran keamanan informasi pada data uji dengan performa yang tinggi, terutama pada kernel *Polynomial*.

### 3. Validasi

Validasi model dilakukan menggunakan data di luar data *training* dan *testing* untuk mengetahui kemampuan model dalam mengklasifikasikan data baru. Hasil validasi disajikan pada Tabel 11:

Tabel 11. Hasil validasi data

| Hasil<br>Klasifikasi | Kernel |                   |            |                |
|----------------------|--------|-------------------|------------|----------------|
|                      | Linear | <i>Polynomial</i> | <i>RBF</i> | <i>Sigmoid</i> |
| Valid                | 23     | 23                | 23         | 10             |
| Tidak Valid          | 0      | 0                 | 0          | 13             |
| Total Data           | 23     | 23                | 23         | 23             |

Berdasarkan hasil evaluasi model klasifikasi empat kernel algoritma *Multi-Class SVM*, diperoleh bahwa kernel *Linear*, *Polynomial*, dan *RBF* menghasilkan performa klasifikasi yang sama, yaitu seluruh data pengujian berhasil diklasifikasikan secara valid sebanyak 23 data dengan tidak terdapat kesalahan klasifikasi. Hal ini menunjukkan bahwa ketiga kernel tersebut mampu mengklasifikasikan kategori tingkat kesadaran keamanan informasi secara konsisten pada data validasi. Sementara itu, kernel *Sigmoid* menghasilkan performa yang lebih rendah dibandingkan kernel lainnya. Dari 23 data pengujian, sebanyak 10 data berhasil diklasifikasikan secara valid, sedangkan 13 data mengalami kesalahan klasifikasi. Perbedaan performa ini menunjukkan bahwa kernel *Sigmoid* kurang mampu menangkap pola pemisahan antar kelas pada dataset penelitian. Berdasarkan hasil evaluasi dan validasi, kernel *Polynomial* dipilih sebagai model terbaik karena menghasilkan performa klasifikasi paling tinggi dan mampu mengklasifikasikan seluruh data validasi dengan benar.

## KESIMPULAN

Penelitian ini berhasil menerapkan algoritma *K-Means* dan *Multi-Class Support Vector Machine* (*SVM*) untuk mengidentifikasi tingkat kesadaran keamanan informasi Pegawai Pemerintah Kota Kendari menggunakan pendekatan *Human Aspects of Information Security Questionnaire* (*HAIS-Q*). Algoritma *K-Means* mampu membentuk tiga kategori tingkat kesadaran berdasarkan karakteristik data responden, sehingga pengelompokan tidak lagi bergantung pada ambang skor yang telah ditentukan sebelumnya. Selain menghasilkan kategori tingkat kesadaran, hasil klusterisasi juga memberikan informasi mengenai kondisi tujuh aspek *HAIS-Q* pada setiap kluster sehingga dapat digunakan sebagai dasar dalam menentukan prioritas peningkatan kesadaran keamanan informasi. Selanjutnya, model *Multi-Class SVM* yang dibangun menggunakan label hasil klusterisasi menunjukkan performa yang sangat baik, di mana kernel *Polynomial* menghasilkan nilai *accuracy*, *precision*, *recall*, dan *F1-score* sebesar 100% serta mampu mengklasifikasikan seluruh data validasi dengan benar. Hasil tersebut menunjukkan bahwa pendekatan yang diusulkan mampu mendukung proses identifikasi tingkat kesadaran keamanan informasi secara efektif berdasarkan karakteristik data responden. Untuk penelitian selanjutnya, penelitian ini dapat dikembangkan dengan melibatkan jumlah responden yang

lebih besar dan lebih beragam, menerapkan algoritma klasterisasi maupun klasifikasi lainnya sebagai pembandingan.

#### DAFTAR PUSTAKA

- Amin, N. F., Garancang, S., & Abunawas, K. (2023). Konsep Umum Populasi dan Sampel Dalam Penelitian. *PILAR: Jurnal Kajian Islam Kontemporer*, 14(1), 15–31. <https://doi.org/10.26618/whw41w62>
- Az-Zahra, A. A., Marsaoly, A. F., Lestyani, I. P., Salsabila, R., & Madjida, W. O. Z. (2021). Penerapan Algoritma K-Modes Clustering dengan Validasi Davies Bouldin Index pada Pengelompokan Tingkat Minat Belanja Online di Provinsi Daerah Istimewa Yogyakarta. *Jurnal Matematika dan Statistika serta Aplikasinya*, 9(1), 24–36. <https://doi.org/10.24252/msa.v9i1.18555>
- Bughin, J., Cincera, M., Reykowska, D., Zyszkiewicz, M., & Ohme, R. (2025). The Great Divide Between Employees: Clustering Employee “Well-Being” During A Pandemic. *PLoS ONE*, 20(3), e0294540. <https://doi.org/10.1371/journal.pone.0294540>
- Dewi, K. M. R., Yudistira, R. D., & Ruldeviyani, Y. (2024). Pengukuran Tingkat Kesadaran Keamanan Informasi Pegawai Pada Instansi Pemerintah. *The Indonesian Journal of Computer Science*, 13(2), 3274–3288. <https://doi.org/10.33022/ijcs.v13i2.3950>
- Fadlika, R., Ruldeviyani, Y., Butarbutar, Z. T., Istiqomah, R. A., & Fariz, A. A. (2023). Employee Information Security Awareness in the Power Generation Sector of PT ABC. *(IJACSA) International Journal of Advanced Computer Science and Applications*, 14(4), 594–603. <https://doi.org/10.14569/IJACSA.2023.014046>
- Gofur, A., Fathoni Aji, R., & Kurniawan, H. (2024). Pengukuran Kesadaran Keamanan Informasi Pegawai: Studi Kasus PT Meshindo Jayatama. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 11(2), 315–320. <https://doi.org/10.25126/jtiik.20241128106>
- Han, H. (2025). Adoption of K-Means Clustering Algorithm in Smart City Security Analysis and Mythical Experience Analysis of Urban Image. *PLoS ONE*, 20(3), e0319620. <https://doi.org/10.1371/journal.pone.0319620>
- Islam, S. S., Haque, M. S., Miah, M. S. U., Sarwar, T. B., & Nugraha, R. (2022). Application of Machine Learning Algorithms to Predict the Thyroid Disease Risk: An Experimental Comparative Study. *PeerJ Computer Science*, 8, e898. <https://doi.org/10.7717/peerj-cs.898>
- Khan, A. R., Khan, S., Harouni, M., Abbasi, R., Iqbal, S., & Mehmood, Z. (2021). Brain Tumor Segmentation using K-Means Clustering and Deep Learning with Synthetic Data Augmentation for Classification. *Microscopy Research and Technique*, 84(7), 1389–1399. <https://doi.org/10.1002/jemt.23694>
- Kusumaningrum, A., Wijayanto, H., & Raharja, B. D. (2022). Pengukuran Tingkat Kesadaran Keamanan Siber di Kalangan Mahasiswa saat Study From Home dengan Multiple Criteria Decision Analysis (MCDA). *Jurnal Ilmiah SINUS*, 20(1), 69–78. <https://doi.org/10.30646/sinus.v20i1.586>
- Liu, G., Mao, S., & Kim, J. H. (2019). A Mature-Tomato Detection Algorithm Using Machine Learning And Color Analysis. *Sensors*, 19(9), 2023. <https://doi.org/10.3390/s19092023>
- Manik, G., Ernawati, I., & Nurlaili, I. (2021). Analisis Sentimen Pada Review Pengguna E-Commerce Bidang Pangan Menggunakan Metode Support Vector Machine (Studi Kasus: Review Sayurbox dan Tanihub pada Google Play). *Prosiding Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA) Jakarta-Indonesia*, 2(2), 64–74.

- Meivira, A., Dewi, N. M. A. R., & Puspitasari, C. E. (2022). Uji Validitas dan Reliabilitas Kuesioner Penggunaan dan Penyimpanan Antibiotika di Kecamatan Ampenan. *Archives Pharmacia*, 4(1), 10–18. <https://doi.org/10.47007/ap.v4i1.4925>
- Monasari, L. D. H., & Pratiwi, N. (2025). Klasifikasi Stunting Pada Balita Menggunakan Algoritma Decision Tree Pada Machine Learning. *Decode: Jurnal Pendidikan Teknologi Informasi*, 5(2), 416–428. <https://doi.org/10.51454/decode.v5i2.1193>
- Nurkholis, A., Alita, D., & Munandar, A. (2022). Comparison of Kernel Support Vector Machine Multi-Class in PPKM Sentiment Analysis on Twitter. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 6(2), 227–233. <https://doi.org/10.29207/resti.v6i2.3906>
- Nurul, S., Anggrainy, S., & Aprelyani, S. (2022). Faktor-faktor yang Mempengaruhi Keamanan Sistem Informasi: Keamanan Informasi, Teknologi Informasi dan Network (Literature Review SIM). *Jurnal Ekonomi Manajemen Sistem Informasi*, 3(5), 564–573. <https://doi.org/10.31933/jemsi.v3i5>
- Prasetya, M. R., Dewi, R. C., Asyhar, A. H., & Hamid, A. (2026). Prediksi Curah Hujan Harian Menggunakan Metode Principal Component Analysis (PCA) dan Long Short-Term Memory (LSTM). *DECODE: Jurnal Pendidikan Teknologi Informasi*, 6(1), 1–13. <https://doi.org/10.51454/decode.v6i1.1451>
- Ramadhan, T., & Purwandari, B. (2023). Analisis Tingkat Kesadaran Keamanan Informasi: Studi Kasus Pengguna Aplikasi Perbankan Digital di Indonesia Guna Mencegah Social Engineering. *Syntax Idea*, 5(1), 86–98. <https://doi.org/10.36418/syntax-idea.v3i6.1227>
- Riyanto, D. A. (2025). *Hootsuite (We are Social): Data Digital Indonesia 2025*. <https://andi.link/hootsuite-we-are-social-data-digital-indonesia-2025/>
- Rohmah, R. N. (2022). Upaya Membangun Kesadaran Keamanan Siber pada Konsumen E-commerce di Indonesia. *Cendekia Niaga Journal of Trade Development and Studies*, 6(1), 1–11. <https://doi.org/10.52391/jcn.v6i1.629>
- Sakalessy, A. J. E., & Purnomo, H. D. (2024). Assessing Employee Performance in the Information Technology Department Using K-Means Clustering: A Case Study Approach. *Journal of Information Systems and Informatics*, 6(1), 170–186. <https://doi.org/10.51519/journalisi.v6i1.653>
- Saridewi, V. S., & Sari, R. F. (2021). Implementation of Machine Learning for Human Aspect in Information Security Awareness. *Journal of Applied Engineering Science*, 19(4), 1126–1142. <https://doi.org/10.5937/jaes0-28530>
- Sofia, A. (2023). K-Means Clustering Using Elbow Method In Case Of Diabetes Milletus Tipe II In Indonesia. *Journal of Statistical Methods and Data Science*, 01(2), 1–8. <https://doi.org/10.31258/jsmds.v1.5>
- Sukman, R. D. W. S. (2024). *Kebocoran Data Nasional Sebanyak 210 Instansi Kena Bobol*. Binus University. <https://sis.binus.ac.id/2024/11/12/kebocoran-data-nasional-sebanyak-210-instansi-kena-bobol/>
- Taufiqurrahman, F., Al Faraby, S., & Purbolaksono, M. D. (2021). Klasifikasi Teks Multi Label pada Hadis Terjemahan Bahasa Indonesia Menggunakan Chi-Square dan SVM. *e-Proceeding of Engineering*, 8(5), 10103–10113. <https://doi.org/10.34818/eoe.v8i5.15671>
- Wu, T., Tien, K. Y., Hsu, W. C., & Wen, F. H. (2021). Assessing The Effects Of Gamification On Enhancing Information Security Awareness Knowledge. *Applied Sciences*, 11(19), 9266. <https://doi.org/10.3390/app11199266>