



PENGELOMPOKAN TINGKAT KESADARAN MASYARAKAT TENTANG PENTINGNYA KEAMANAN INFORMASI DATA PRIBADI BERDASARKAN CLUSTERISASI K-MEANS MENGUNAKAN EUCLIDIEN DISTANCE

Yoyon Arie Budi Suprio¹, Mohamad Farid²

^{1,2} Sekolah Tinggi Ilmu Komputer PGRI Banyuwangi

Jl. Jend A. Yani 80, Banyuwangi Kode Pos 68416

e-mail : yoyonstikom@gmail.com¹, faridmuhamad2001@gmail.com²

ABSTRAK

Keamanan informasi adalah salah satu hal penting yang harus diperhatikan oleh semua orang. Banyaknya celah bocornya keamanan informasi menyebabkan semua orang harus berhati-hati dalam berbagai hal. Bocornya keamanan informasi dapat menyebabkan risiko yang sangat fatal, diantaranya penggunaan data pribadi sembarangan dan lain sebagainya. Berdasarkan permasalahan tersebut dilakukan pengukuran masalah dari dimensi kesadaran (attitude, knowledge dan behaviour) dengan menggunakan metode clustering untuk penentuan kelompok kesadaran masyarakat. Dengan algoritma K-Means sistem akan mengolah data dan mengelompokkan sesuai dengan nilai terdekat dengan titik pusat cluster. Dengan adanya penelitian ini akan didapat kelompok kesadaran masyarakat dengan tingkat rendah, sedang, dan tinggi sehingga dapat dilakukan pembinaan pentingnya kesadaran masyarakat tentang keamanan informasi mereka sesuai dengan kelompok yang didapat.

Kata kunci : *Clustering, k-means, keamanan informasi*

ABSTRACT

Information security is one of the important things that everyone should pay attention to. The number of leaking information security loopholes causes everyone to be careful in various matters. Leaking information security can cause very fatal risks, including the use of personal data indiscriminately and so on. Based on these problems, measurement of problems from the dimension of awareness (attitude, knowledge and behavior) by using clustering methods for the determination of community awareness groups. With the K-Means algorithm the system will process the data and group according to the value closest to the center point of the cluster. With this research will be obtained by public awareness groups with low, medium, and high levels so that public awareness can be done in accordance with the group obtained.

Kata kunci : *Clustering, k-mean, information security*

1. PENDAHULUAN

Keamanan informasi adalah melindungi informasi dari berbagai ancaman demi menjamin kelangsungan proses bisnis, meminimalisir risiko bisnis dan memaksimalkan laba investasi dan peluang bisnis. Keamanan informasi dapat dibentuk dengan cara menerapkan suatu set kontrol yang termasuk didalamnya kebijakan, proses, prosedur, struktur organisasi serta fungsi dari perangkat lunak dan perangkat keras.

Kontrol tersebut perlu ditetapkan, dilaksanakan, dipantau, dikaji ulang dan disempurnakan demi menjamin keamanan dan tercapainya tujuan bisnis organisasi. (Syafitri et al., 2016)

Di era semakin pesatnya perkembangan teknologi seperti saat ini, banyak ancaman yang dapat terjadi. Salah satunya yaitu ancaman bocornya keamanan informasi data pribadi. Bocornya keamanan informasi terjadi karena faktor internal dan eksternal. (Hasanah et al.,



2017) Kebocoran keamanan informasi dapat menyebabkan risiko yang sangat fatal. Meningkatnya tren ancaman keamanan informasi di dunia, misalnya serangan siber semakin canggih dan masif, ransomware menyerang setiap 14 detik pada tahun 2019 (sumber: <https://techjury.net/>).

Belakangan ini kasus kebocoran informasi marak terjadi, sehingga evaluasi keamanan informasi sangat diperlukan agar bisa memberikan gambaran tentang kesiapan dalam menghadapi hal tersebut.(Sni et al., 2021). Oleh karena itu kita harus memperhatikan beberapa faktor baik internal maupun eksternal.

Faktor internal adalah penyebab terbesar bocornya keamanan informasi. Kurangnya edukasi tentang celah bocornya keamanan informasi menyebabkan masyarakat tidak mengathui tentang ancaman tersebut. Maka dari itu, perlu adanya pembinaan tentang pentingnya keamanan informasi.

Penelitian ini diharapkan dapat menghasilkan data kelompok persebaran masyarakat menjadi 3 (rendah, sedang, dan tinggi), sehingga dapat dilakukan pembinaan sesuai kelompok tersebut.

Penentuan kelompok ini dilakukan berdasarkan data survey yang didapat. Data survey bisa dijadikan sebagai cerminan sadarnya masyarakat tentang kemanan informasi mereka.

Penentuan kelompok masyarakat tentang pentingnya keamanan informasi menggunakan *clustering* dengan algoritma *k-means*. Algoritma ini cocok untuk penelitian ini dikarenakan dengan mencari cluster terdekat dengan titik pusat cluster sehingga bisa menemukan persebaran masyarakat. Hal tersebut bisa dijadikan acuan untuk penentuan kelompok berdasarkan kesadaran masyarakat.

Dengan adanya penelitian ini diharapkan dapat memberikan solusi untuk permasalahan keamanan informasi. Sehingga masyarakat sadar akan pentingnya keamanan informasi mereka dan tidak ditemukan lagi gangguan keamanan informasi yang terjadi di masyarakat.

Algoritma k-means

Ada berbagai algoritma yang dapat digunakan untuk menentukan kelompok, salah

satunya adalah *K-Means*. Pengertian dari *K-Means Clustering* adalah, *K* dimaksudkan sebagai konstanta jumlah cluster yang diinginkan, Means dalam hal ini berarti nilai suatu rata-rata dari suatu grup data yang dalam hal ini didefinisikan sebagai cluster.

Metode *K-Means* berusaha mengelompokkan data yang ada kedalam beberapa kelompok, dimana data dalam satu kelompok mempunyai karakteristik yang sama satu sama lainnya dan mempunyai karakteristik yang berbeda dengan data yang ada didalam kelompok yang lain.(Manihuruk et al., 2020)

Perhitungan *k-means* menggunakan *eucliden distance* sebagai berikut:

1. Tentukan nilai *k* sebagai jumlah klaster yang akan dibentuk.
2. Umpamakan *k* sebagai centroid yang dapat ditentukan secara acak.
3. Hitung jarak setiap data ke centroid menggunakan persamaan *Euclidean Distance*.
4. Kelompokkan setiap data berdasarkan jarak terdekat antara data dengan centroidnya.
5. Tentukan posisi centroid baru (*k*)
6. Ulangi langkah 3 jika letak centroid baru dan lama tidak sama.

Dasar Algoritma K-Means

Algoritma *K-Means* adalah bekerja dengan membagi data ke dalam *k* buah cluster yang telah ditentukan.(Agustina & Prihandoko, 2018)

Beberapa cara penghitungan jarak yang biasa digunakan, salah satunya yaitu *Eucliden Distance*. *Euclidean distance* adalah perhitungan untuk mengukur jarak dua titik dalam *euclidean space* yang mempelajari hubungan antara sudut dan jarak.(Miftahuddin et al., 2020) Untuk mengukur tingkat kemiripan data dengan rumus *euclidean distance* digunakan rumus berikut:

$$d(x, y) = |x - y| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

Keterangan:

- d : jarak antara x dan y
- x : data pusat klaster
- y : data pada atribut

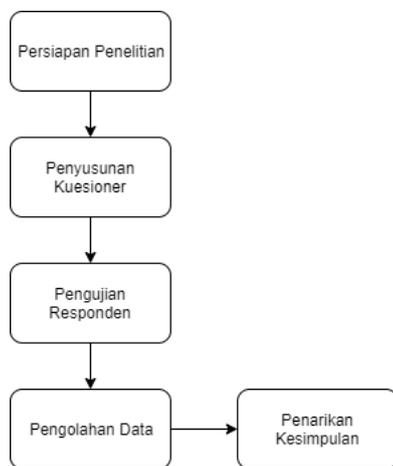


- i : setiap data
- n : jumlah data
- xi : data pada pusat klaster ke i
- yi : data pada setiap data ke i

2. METODOLOGI PENELITIAN

Metode penelitian ini menggunakan clusterisasi dengan algoritma *k-means*. Clusterisasi merupakan aktivitas yang bertujuan melakukan pengelompokan data yang memiliki kemiripan antara satu data dengan data lainnya ke dalam klaster atau kelompok sehingga data dalam satu klaster memiliki tingkat kemiripan. *Clustering* juga dapat diartikan metode segmentasi data yang diimplementasikan dalam beberapa bidang, diantaranya marketing, analisa masalah bisnis segmentasi pasar dan prediksi, pola dalam bidang computer vision, zonasi wilayah hingga identifikasi obyek dan pengolahan citra. Analisis klaster bertujuan menemukan kelompok objek sedemikian rupa sehingga objek-objek dalam grup akan sama (atau terkait) satu sama lain dan berbeda dari (atau tidak terkait) objek-objek dalam grup lain. (Nishom, 2019)

Penelitian yang digunakan adalah penelitian kuantitatif.



Gambar 1. Tahapan Penelitian

Berikut adalah Tahapan Penelitian yang dilakukan.

- a. Persiapan Penelitian

Pada tahap ini peneliti melakukan pendalaman studi kasus yang akan dilakukan, kemudian menentukan solusi apa yang cocok. Pada tahap ini peneliti juga mencari referensi sebagai pendukung dan penguatan materi untuk pengerjaan penelitian ini.
- b. Penyusunan Kuesioner

Penyusunan kuesioner dilakukan berdasarkan *attitude, knowledge, dan behavior* responden guna mengetahui kesadaran tentang pentingnya keamanan informasi data pribadi. Penelitian ini memiliki 9 pertanyaan yang dijawab dalam skala 3 pilihan yaitu sangat penting, penting, tidak penting, untuk menguji *attitude, knowledge dan behavior* mereka. Pertanyaan yang diajukan dapat dilihat di Tabel 1. Kuesioner disebar secara online.

Tabel 1. Tabel Pertanyaan

Poin	Pertanyaan	Jawaban
Knowledge	- penggunaan password atau pin yang sama untuk beberapa akun dan aplikasi adalah hal yang harus dihindari	1. Sangat Penting 2. Penting 3. Tidak Penting
	- pentingnya meninggalkan smartphone di tempat yang aman	
	- membiarkan orang lain mengetahui password atau pin adalah hal yang harus dihindari	
Attitude	- saya sadar pentingnya	



Behavior	<ul style="list-style-type: none"> - untuk menyimpan password atau pin hanya untuk diri sendiri - saya sadar pentingnya menggunakan password atau pin yang komplek agar tidak mudah ditebak - pentingnya menyimpan OTP hanya untuk diri sendiri - pentingnya menginstal aplikasi hanya dari Google Play Store (Android) atau Apple App Store (IOS) - pentingnya tidak membuka aplikasi yang mengandung password ketika terhubung ke jaringan WIFI umum - pentingnya melakukan pengamanan menggunakan lockscreen di smartphome, baik dengan password, PIN, pola atau yang lainnya
-----------------	--

- c. Pengujian Responden
Pengujian responden ini dilakukan melalui pengisian kuesioner yang peneliti berikan kepada responden.
- d. Pengolahan Data
Setelah dilakukan pengujian responden, tahap selanjutnya yaitu melakukan

pengolahan data berdasarkan hasil pengisian kuesioner oleh responden.

e. Penarikan Kesimpulan
Penarikan kesimpulan dilakukan jika pengolahan data sudah selesai dilakukan. Di tahap ini akan diketahui apakah penelitian ini sesuai dengan tujuan yang dihapkan atau tidak.

3. HASIL DAN PEMBAHASAN

Dalam penelitian ini mengambil 20 sampel responden yang didapat dengan mengisikan kuesioner yang telah diberikan. Responden terdiri dari berbagai lapisan masyarakat sehingga dapat dijadikan acuan untuk penentuan kelompok kesadaran masyarakat. Di bawah ini merupakan karakteristik dari responden yang ada.

Tabel 2. Tabel pengalaman gangguan keamanan informasi

No.	Pengalaman Gangguan Keamanan Informasi	Presentase
1.	Ya	95%
2.	Tidak	5%

Dari tabel diatas diketahui bahwa 95% responden pernah mengalami gangguan keamanan informasi.

Tabel 3. Tabel Jenis Kelamin Responden

No.	Jenis Kelamin	Presentase
1.	Laki-laki	45%
2.	Perempuan	55%

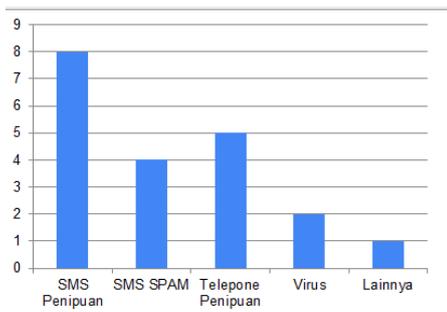
Berdasarkan tabel diatas menunjukkan bahwa responden perempuan lebih banyak dari responden laki-laki. Kemudian diketahui juga apakah responden pernah mengalami gangguan keamanan informasi.

Tabel 4. Tabel Gangguan Keamanan Informasi

No.	Terganggu dengan Adanya Gangguan Keamanan Informasi	Presentase
1.	Ya	95%
2.	Tidak	5%



Berdasarkan tabel diatas dapat diketahui bahwa 95% responden pernah mengalami gangguan keamanan informasi. Gangguan keamanan informasi bisa berasal dari mana saja, berikut adalah tabel bentuk gangguan keamanan informasi



Gambar 2. Bentuk Gangguan Keamanan Informasi

Gambar diatas menunjukkan bahwa gangguan keamanan informasi yang banyak didapat adalah dari SMS Penipuan.

Berikut ini adalah pembahasan dari hasil penelitian yaitu data yang didapat dari kuisioner:

No	P1	P2	P3	P4	P5	P6	P7	P8	P9
1	1	1	2	2	1	1	2	2	1
2	1	1	1	1	1	1	1	1	1
3	3	1	1	1	1	1	1	1	1
4	1	1	3	1	1	1	1	1	1
5	1	1	2	2	2	2	1	2	1
6	1	1	2	1	2	1	1	3	1
7	2	2	2	2	2	2	2	2	2
8	3	3	3	3	3	3	3	3	2
9	2	1	1	2	1	2	2	2	2
10	2	1	3	1	1	1	1	1	1
11	1	1	1	2	2	2	2	2	1
12	2	1	1	2	1	2	3	3	2
13	2	3	3	3	3	3	3	2	3
14	2	1	1	1	2	2	2	2	2
15	1	3	2	2	2	2	2	2	2
16	2	1	1	1	1	1	2	2	2
17	1	1	1	1	1	1	1	2	2
18	2	2	1	2	1	2	2	2	2
19	3	1	3	1	1	1	1	1	1
20	2	1	1	1	1	1	2	2	1

Gambar 3. Data kuisioner

Maka langkah selanjutnya yaitu menentukan pusat cluster secara random, dan mengambil data sebagai berikut:

Iterasi I										
Cluster	No	P1	P2	P3	P4	P5	P6	P7	P8	P9
C1	1	1	1	2	2	1	1	2	2	1
C2	4	1	1	3	1	1	1	1	1	1
C3	9	2	1	1	2	1	2	2	2	2

Gambar 4. Pusat Cluster 1

Kemudian dilakukan perhitungan Iterasi I dengan rumus *Eucliden Distance* dan menghasilkan data sebagai berikut:

Data	C1	C2	C3	Keanggotaan	Min Jarak	Min Jarak*2
1	0	2,236068	1,732051	C1	0	0
2	2	2	2,44949	C1	2	4
3	2,828427	2,828427	2,44949	C3	2,44949	6
4	2	0	3,162278	C2	0	0
5	1,414214	2	2,44949	C1	1,414214	2
6	2	2,44949	2,828427	C1	2	4
7	2,236068	3	1,732051	C3	1,732051	3
8	4,582576	5,385165	4,123106	C3	4,123106	17
9	2	3,162278	0	C3	0	0
10	2,236068	1	3	C2	1	1
11	1,732051	3	1,732051	C1	1,732051	3
12	3	4,358899	1,732051	C3	1,732051	3
13	4,472136	5,09902	4	C3	4	16
14	2,44949	3,162278	1,414214	C3	1,414214	2
15	2,645751	3,316625	2,645751	C1	2,645751	7
16	2,236068	3	1	C3	1	1
17	2	2,44949	2	C1	2	4
18	2,236068	3,316625	1	C3	1	1
19	2,828427	2	3,162278	C2	2	4
20	2	2,828427	1,414214	C3	1,414214	2
				WCV		80

Gambar 5. Hasil Iterasi 1

Langkah selanjutnya yaitu menghitung jarak antara pusat Cluster dengan rumus *Eucliden Distance* dan menghasilkan data sebagai berikut:

Hubungan	Jarak
C1 C2	2
C1 C3	2
C2 C3	3,162278
BCV	7,162278

Gambar 6. Hasil Perhitungan Jarak Pusat Cluster 1

Lalu menghitung Rasio Iterasi I dengan rumus

$$Rasio = \frac{BCV}{WCV} = \frac{7,162278}{80} = 0,089528471 \quad (2)$$

Kemudian menghitung Iterasi II, yang pertama dilakukan adalah menentukan pusat cluster secara random, dan saya mengambil data sebagai berikut:

Iterasi II										
Cluster	No	P1	P2	P3	P4	P5	P6	P7	P8	P9
C1	8	3	3	3	3	3	3	3	3	2
C2	13	2	3	3	3	3	3	3	2	3
C3	19	3	1	3	1	1	1	1	1	1

Gambar 7. Pusat Cluster 2

Kemudian dilakukan perhitungan Iterasi II dengan rumus *Eucliden Distance* dan menghasilkan data sebagai berikut:



Data	C1	C2	C3	Keanggotaan	Min Jarak	Min Jarak*	
1	4,582576	4,472136	2,828427	C3	2,828427		
2	5,744563	5,477226	2,828427	C3	2,828427		
3	5,385165	5,477226		2 C3		2	
4	5,385165	5,09902		2 C3		2	
5	4,242641	4,123106		3 C3		3	
6	4,795832	4,898979	3,162278	C3	3,162278	1	
7	2,828427	2,645751		3 C3	2,645751		
8	0	1,732051		5 C1		0	
9	4,123106		3,162278	C3	3,162278	1	
10	5,09902			1 C3		1	
11	4,242641	4,123106	3,605551	C3	3,605551	1	
12	3,872983		4	4 C1	3,872983	1	
13	1,732051		5,09902	C3		0	
14	4,123106		4	3,162278	C3	3,162278	1
15	3,162278	2,645751	3,872983	C3	2,645751		
16	4,795832	4,690416	2,828427	C3	2,828427		
17	5,385165	5,09902	3,162278	C3	3,162278	1	
18	3,741657	3,605551	3,316625	C3	3,316625	1	
19		5,09902		0 C3		0	
20	4,898979		5,2645751	C3	2,645751		
				WCV		14	

Gambar 8. Hasil Iterasi 2

Langkah selanjutnya yaitu menghitung jarak antara pusat Cluster dengan rumus *Eucliden Distance* dan menghasilkan data sebagai berikut:

Hubungan		Jarak
C1	C2	1,732051
C1	C3	5
C2	C3	5,09902
BCV		11,83107

Gambar 9. Hasil Perhitungan Jarak Pusat Cluster 2

Rasio = $11,83107/142$
 = $0,083317397$

Rasio Iterasi I : $0,089528471$
 Rasio Iterasi II : $0,083317397$

Perhitungan Iterasi akan selesai apabila nilai rasio saat ini lebih kecil atau sama dengan nilai rasio iterasi sebelumnya. Karena nilai rasio iterasi II ($0,083317397$) lebih kecil dari nilai rasio iterasi I ($0,089528471$), maka proses iterasi selesai dan diketahui Keanggotaannya sebagai berikut:

- C1 : 8, 10, dan 12.
- C2 : 7, 13, dan 15.
- C3 : 1, 2, 3, 4, 5, 6, 9, 10, 11, 14, 16, 17, 18, 19, dan 20.

Tingkat Kesadaran Informasi

- C1 : Sedang
- C2 : Rendah
- C3 : Tinggi

Presentase Keanggotaan tingkat kesadaran masyarakat tentang keamanan informasi.

Presentase Tingkat Kesadaran Masyarakat



Gambar 10. Persentase Tingkat Kesadaran Masyarakat

4. KESIMPULAN

Dengan keamanan informasi ini maka melindungi informasi dari berbagai ancaman. Kesadaran masyarakat tentang pentingnya keamanan informasi mereka dapat diketahui dari 20 sampel responden yang telah mengisi kuisioner, didapat 95% dari mereka pernah mengalami gangguan keamanan informasi. Setelah dilakukan clusterisasi menggunakan *k-means* didapat 3 kelompok masyarakat. Kelompok 1 dengan tingkat kesadaran sedang memiliki presentase 15% dari total responden. Kelompok 2 dengan tingkat kesadaran rendah memiliki presentase 15%. Sedangkan kelompok 3 dengan tingkat kesadaran tinggi memiliki presentase 70%.

Dari hasil penelitian tersebut 70% responden memiliki tingkat kesadaran tinggi. Sehingga disimpulkan bahwa masyarakat sadar akan pentingnya keamanan informasi akan tetapi masih mengalami gangguan keamanan informasi.

Setelah didapat 3 kelompok tersebut diharapkan dapat dilakukan pembinaan tentang pentingnya keamanan informasi sesuai tingkat kesadaran mereka, sehingga tidak akan terjadi lagi gangguan keamanan informasi pada mereka.

5. REFERENSI

Agustina, N., & Prihandoko, P. (2018). Perbandingan Algoritma K-Means dengan Fuzzy C-Means Untuk Clustering Tingkat Kedisiplinan Kinerja Karyawan. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 2(3), 621–626. <https://doi.org/10.29207/resti.v2i3.492>



Hasanah, Y. M., Safruddin, C., Jabar, A., Pauh, T., Yogyakarta, U. N., Hasanah, Y. M., Safruddin, C., Jabar, A., Pauh, T., & Yogyakarta, U. N. (2017). Jurnal Akuntabilitas Manajemen Pendidikan Online :

<http://journal.uny.ac.id/index.php/jamp>

Jurnal Akuntabilitas Manajemen Pendidikan Pendahuluan Pendidikan merupakan hal paling penting dan investasi masa yang akan datang dalam suatu negara . Pendidikan. *Jurnal Akuntabilitas Manajemen Pendidikan*, 5(2), 228–239.

Manihuruk, N. A., Zarlis, M., Irawan, E., & Tambunan, H. S. (2020). Penerapan Data Mining Dalam Mengelompokkan Calon Penerima Beasiswa Dengan Menggunakan Algoritma K-Means. *KOMIK (Konferensi Nasional Teknologi Informasi Dan Komputer)*, 4(1), 29–34. <https://doi.org/10.30865/komik.v4i1.2575>

Miftahuddin, Y., Umaroh, S., & Karim, F. R. (2020). Perbandingan Metode Perhitungan Jarak Euclidean, Haversine, Dan Manhattan Dalam Penentuan Posisi Karyawan. *Jurnal Tekno Insentif*, 14(2), 69–77. <https://doi.org/10.36787/jti.v14i2.270>

Nishom, M. (2019). Perbandingan Akurasi Euclidean Distance, Minkowski Distance, dan Manhattan Distance pada Algoritma K-Means Clustering berbasis Chi-Square. *Jurnal Informatika: Jurnal Pengembangan IT*, 4(1), 20–24. <https://doi.org/10.30591/jpit.v4i1.1253>

Sni, B. I. S. O., Rochmadi, T., & Pasa, I. Y. (2021). *Menggunakan Indeks Keamanan Informasi Di Bkd Xyz Measurement of Risk and Evaluation of Information Security Using the Information Security Index in Bkd Xyz Based on Iso 27001 / Sni*. 4(1), 38–43.

Syafitri, W., Akademik, S. I., Xyz, U., Akademik, S. I., & Xyz, U. (2016). *Penilaian Risiko Keamanan Informasi Menggunakan Metode NIST 800-30 (Studi Kasus : Sistem Informasi Akademik Universitas XYZ)*. 2(2), 8–13.