p-ISSN: 2252-3006
e-ISSN: 2685-2411

Analisis Data Log Honeypot Menggunakan Metode K-Means Clustering

Gede Haris Premana Wibawa, I Gusti Made Arya Sasmita, I Made Sunia Raharja
Program Studi Teknologi Informasi, Fakultas Teknik, Universitas Udayana
Bukit Jimbaran, Bali, Indonesia, telp. (0361) 701806
e-mail: harisfertik@gmail.com aryasasmita88@gmail.com sunia.raharja@gmail.com

Abstrak

Perkembangan teknologi informasi saat ini begitu pesat termasuk pada jaringan komputer, menuntut agar sistem keamanan untuk berkembang. Honeypot adalah sebuah sistem yang didesain mirip dengan sistem yang asli dan dibuat dengan maksud untuk diserang sehingga sistem asli masih tetap aman. Honeypot sendiri mendeteksi serangan dan mencatat serangan dalam bentuk data log. Dari data log tersebut dapat digunakan untuk menganalisis sebuah serangan karena data log menyimpan data penyerang mulai dari IP address, port yang diserang dan layanan yang diserang. Dari data log yang disimpan dengan pengambilan data selama 30 hari. Penelitian ini dilakukan bertujuan untuk mengetahui kerentanan serangan pada Server salah satu Institusi di Bali. Metode K-Means digunakan untuk mengelompokan data serangan. Hasil dari K-Means Clustering pada data log memperoleh tiga cluster, dua diantaranya berada pada kategori risiko Low dan satu cluster masuk pada kategori risiko Medium dengan nilai silhouette score adalah 0.999.

Kata kunci: *Data Log, Data Mining, Honeypot K-Means Clustering*

Abstract

The development of information technology is currently increasing, including in computer networks, demanding that the security system to evolve. Honeypot is a system that is designed similar to the original system and is made with the intention to be attacked so that the original system is still safe. The Honeypot logs themselves are attacks and log attacks are recorded. This log data can be used to analyze attacks because log data stores the attacker's data starting from the IP address, the port being attacked and the service being attacked. From the log data that is stored by retrieving data for 30 days. This research was conducted to find out how to deal with one of the Institutions in Bali Server. The K-Means method is used to classify data attacks. The results of K-Means Clustering in the log data obtained by three clusters, two that passed in the Low risk category and one cluster included in the Medium risk category with a silhouette score is 0.999.

Keywords: *Data Log, Data Mining, Honeypot K-Means Clustering*

1. Pendahuluan

Perkembangan teknologi informasi saat ini begitu pesat ditandai dengan semua kegiatan menjadi lebih mudah termasuk pada jaringan komputer [1]. Jaringan komputer terhubung melalui internet sebagai penghubung utama dalam pertukaran data maupun informasi. Internet merupakan jaringan yang bersifat publik dan terjangkau luas, oleh sebab itu diperlukan adanya upaya untuk menjaga dan menjamin keamanan informasi terhadap layanan yang berada di internet. Banyak strategi dalam menjaga keamanan informasi seperti penerapan *firewall* atau *Intrusion Detection System* (IDS) yang mampu menjaga keamanan layanan, tetapi masih ada banyak masalah yang ditemukan [2]. Sistem peringatan IDS dengan antarmuka berbasis *website* seperti BASE tidak dapat menawarkan pemberitahuan kepada administrator, ada kemungkinan bahwa pengguna mungkin kehilangan beberapa serangan sehingga respons menjadi terlambat untuk dilakukan [3].

Honeypot merupakan sebuah teknologi untuk melindungi aset dari penyalahgunaan informasi yang disebabkan dari kejahatan siber dalam beberapa tahun terakhir [4]. *Honeypot* juga dapat digunakan untuk meningkatkan deteksi keamanan perusahaan. Cara *honeypot* ialah

p-ISSN: 2252-3006
e-ISSN: 2685-2411

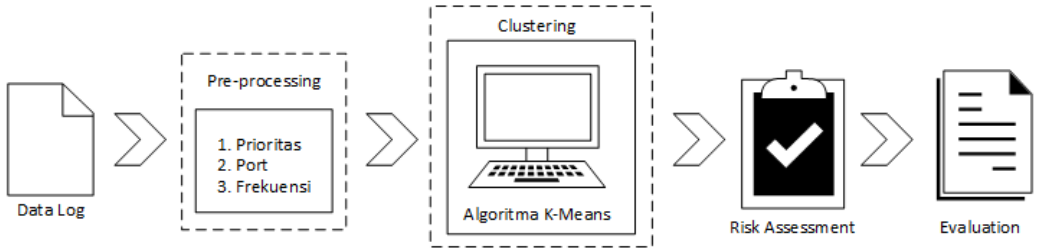
memancing penyerang agar berinteraksi dan mengumpulkan informasi yang akan digunakan untuk dianalisis [5]. Berdasarkan tingkat interaksi *honeypot* dibedakan menjadi tiga, diantaranya *low-interaction honyepots*, *medium-interaction*, dan *high-interaction* [6]. Selain itu *honeypot* dapat mendeteksi dan menyimpan informasi serangan yang tersimpan dalam bentuk *data log*. Informasi yang disimpan pada *data log* meliputi data penyerang berupa IP *address*, *port* yang diserang, layanan yang diserang dan waktu penyerangan. Pengelompokan data serangan yang tercatat dalam bentuk *data log* merupakan solusi yang tepat untuk menentukan tingkat serangan yang dapat dideteksi. Pengelompokan data serangan memudahkan administrator *server* memperoleh informasi yang mudah dipahami dibandingkan harus membaca *data log* yang sulit untuk dipahami. Ada banyak algoritma yang dapat diterapkan dalam melakukan pengelompokkan data diantaranya seperti Fuzzy C-Means (FCM), DBSCAN, *Hierarchical Clustering*, dan K-Means *Clustering*. Beberapa penelitian yang sudah dilakukan dalam menerapkan metode K-Means Clustering diantaranya, penggunaan K-Means *clustering* pada *log file* untuk medeteksi anomali trafik [7]. Penelitian tersebut menggunakan metode K-Means *Clustering* untuk memberikan informasi berdasarkan hasil dari pendeteksian *log* dengan menentukan pola yang ada. Hasil dari analisa *data log* berdasarkan *cluster* akan terbentuk pola trafik data serangan yang dilakukan oleh penyerang dengan akses jaringan lain. Penelitian lainnya mengenai pemanfaatan metode K-Means *Clustering* untuk mengenali perilaku pengguna internet berdasarkan *data log* jaringan di Lembaga Pendidikan salah universitas di Yogyakarta [8]. Penelitian tersebut menggunakan metode K-Means untuk pengelompokkan berdasarkan jumlah pengunjung dengan jumlah *cluster* dibagi menjadi tiga. Hasil pengelompokkan menunjukkan bahwa atribut data berdasarkan waktu akses pagi-sore dan sore-malam memiliki pengunjung terbanyak.

Penulis memilih metode K-Means *Clustering* sebagai metode yang digunakan untuk pengelompokkan *data log honeypot* dan penentuan kerentanan serangan pada *Server* di Institusi X. Perbedaan penelitian ini dari penelitian sebelumnya terletak pada data yang digunakan yaitu *data log honeypot* yang terpasang pada *Server* salah satu Institusi di Bali dan perhitungan nilai risiko. Sebelum dilakukan proses *clustering*, *data log* akan dilakukan penambahan tiga parameter untuk mempermudah proses *clustering*. Kemudian dilakukan perhitungan nilai risiko serangan dari hasil *clustering* dimana kategori risiko dibedakan menjadi tiga yaitu serangan *Low*, serangan *Medium*, dan serangan *High*.

2. Metodologi Penelitian

Pengumpulan data pada penelitian ini menggunakan metodologi observasi langsung dengan pemasangan sistem *honeypot* pada *Server* salah satu Institusi di Bali. *Honeypot* mendeteksi serangan melalui trafik yang masuk ke *Server* dari Insitusi X selama 30 hari dan hasil deteksi serangan tersebut disimpan dalam sebuah *data log*.

Metode yang digunakan dalam pengelompokkan *data log honeypot* menggunakan metode K-Means *Clustering*, kemudian menentukan kategori risiko serangan dari setiap data melalui perhitungan nilai risiko serangan. Tahap *preprocessing* dilakukan untuk mempermudah proses *clustering* dengan menambahkan tiga parameter dari setiap data serangan pada penelitian ini diantaranya Prioritas, *Port*, dan Frekuensi. Gambaran umum penelitian ini dapat dilihat pada Gambar 1.



Gambar 1. Gambaran Umum

2.1 Preprocessing

Tahap *preprocessing* bertujuan untuk penambahan tiga parameter yang digunakan untuk memenuhi syarat tingkatan risiko dari suatu data serangan, dimana tiga parameter

p-ISSN: 2252-3006
e-ISSN: 2685-2411

tersebut diantaranya Prioritas, *Port*, dan Frekuensi. Adapun penjelasan tiga parameter tersebut dapat dilihat pada Tabel 1.

Tabel 1. Parameter Nilai Risiko

Parameter	Deskripsi
Prioritas	Prioritas menentukan urutan dari serangan yang harus diwaspadai. Pengelompokan prioritas dibedakan menjadi tiga diantaranya *Low* = 1-2, *Medium* = 3, dan *High* = 4.
Port	*Port* merupakan nilai untuk menentukan pentingnya suatu *Port*. Pengelompokan *port* dibedakan menjadi tiga diantaranya *Well-Known Port* (0-1023) = 4, *Registered Port* (1024-49151) = 3, dan *Dynamic / Private Port* (45152-65535) = 1-2.
Frekuensi	Parameter ini didefinisikan untuk frekuensi dari setiap jenis serangan. Nilai parameter ini adalah antara 1-2 = *Low*, 3-4 = *Medium*, dan 5 = *High* [9].

2.2 Clustering

Tahap *clustering* merupakan pengelompokkan data serangan dari *data log* yang telah melalui tahap *preprocessing*, dimana data serangan yang memiliki karaterisktik berdasarkan tiga parameter tersebut menggunakan metode K-Means *Clustering*. Penentuan jumlah *cluster* yang paling optimal berdasarkan data serangan yang terdiri dari 308.781 data ialah berjumlah 3 *cluster*.

2.3 Risk Assessment

Tahap *risk assessment* merupakan tahap dimana hasil *clustering* menghasilkan tiga titik pusat *centroid* akhir yang membedakan dari ketiga *cluster* tersebut. Adapun perhitungan nilai risiko serangan dihitung berdasarkan tiga titik pusat *centroid* akhir dari hasil *clustering* [10]. Perhitungan nilai risiko berdasarkan tiga parameter dapat dilihat pada Tabel 1, dapat menggunakan persamaan sebagai berikut.

Prioritas (*P*) = {1-4}
Port (*D*) = {1-4}
Frekuensi (*F*) = {1-5}
MaxRA= 10

$$RA = \frac{P * D * F}{X} \quad (1)$$

Parameter Prioritas (*P*) memiliki rentang nilai diantara 1 hingga 4, parameter *Port* (*D*) memilki rentang nilai diantara 1 hingga 4, parameter Frekuensi (*F*) memiliki rentang nilai diantara 1 hingga 5. Maksimal nilai risiko memiliki rentang nilai diantara 1 hingga 10 dan nilai *X* dapat diperoleh dengan persamaan sebagai berikut.

Max(*P*)=4, Max(*D*)=4, Max(*F*)=5

$$RA = \frac{4 * 4 * 5}{X} = 10 \quad (2)$$

$$X = \frac{80}{10} = 8 \quad (3)$$

$$RA = \frac{P * D * F}{8} \quad (4)$$

Maksimal nilai parameter (*P*) adalah 4, maksimal nilai parameter (*D*) adalah 4, dan maksimal nilai parameter (*F*) adalah 5. Nilai *X* yang diperoleh adalah 8 berdasarkan hasil

p-ISSN: 2252-3006
e-ISSN: 2685-2411

dari Persamaan 3. Fungsi *RA* digunakan untuk mencari nilai risiko berdasarkan titik pusat *centroid* akhir pada masing-masing *cluster*. Hasil dari nilai risiko yang telah diperoleh dapat dipetakan berdasarkan kategori risiko serangan yang dapat dilihat pada Tabel 2.

Tabel 2. Kategori Risiko Serangan

Nilai Risiko	Kategori
1-4	Low
5-7	Medium
8-10	High

Tabel 2 merupakan kategori risiko serangan berdasarkan hasil dari nilai risiko. Hasil nilai risiko dengan rentang nilai 1 hingga 4 termasuk dalam kategori risiko serangan *Low*, hasil nilai risiko dengan rentang nilai 5 hingga 7 masuk dalam kategori risiko serangan *Medium*, dan hasil nilai irisko dengan rentang nilai 8 hingga 10 masuk kedalam kategori risiko serangan *High*.

2.4 Evaluation

Hasil *clustering* perlu dilakukan evaluasi untuk menentukan bahwa hasil yang telah diperoleh sudah optimal atau tidak. Ada beberapa metode yang digunakan salah satunya metode *Silhouette Coefficient*. Metode ini digunakan untuk mengetahui kualitas dari hasil *clustering* metode K-Means.

3. Kajian Pustaka

Kajian pustaka pada penelitian ini memuat tentang studi literatur yang menjadi referensi penelitian ini. Beberapa materi yang dimuat diantaranya adalah *Honeypot*, *Data Mining*, metode K-Means *Clustering*, dan metode *Silhouette Coefficient*.

3.1 Honeypot

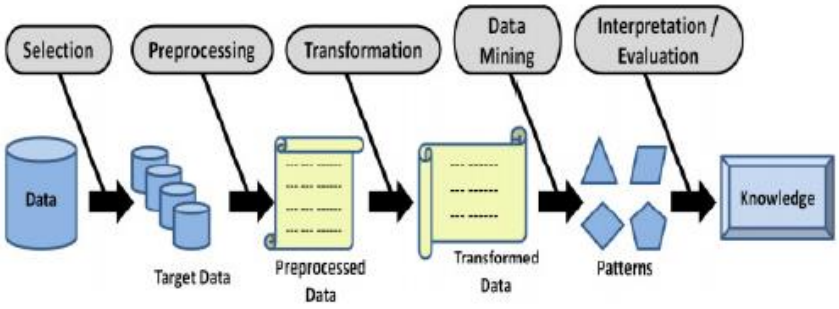
Honeypot merupakan sebuah sistem palsu atau layanan palsu yang berperan sebagai penjebak pengguna yang memiliki maksud buruk atau menangkal upaya yang dapat merugikan layanan [11]. *Honeypot* berfungsi sebagai pengalih perhatian penyerang, agar penyerang merasa berhasil menyusup dan mengambil informasi dari jaringan, namun sebenarnya informasi tersebut tidak penting dan lokasi tersebut sudah terisolir.

Honeypot juga bermanfaat untuk para *administrator* dalam menganalisa aktivitas yang dilakukan oleh penyerang ketika menyusup sistem *honeypot* [12]. Sistem *honeypot* mampu menyimpan banyak data serangan dan melindungi *server* asli sehingga *server* asli masih bisa berjalan dengan baik. Data dari hasil serangan akan disimpan dalam sebuah *data log* yang dapat dianalisa dan menghasilkan informasi dari analisa tersebut.

3.2 Data Mining

Data mining merupakan sebuah proses pencarian secara otomatis informasi yang berguna dalam tempat penyimpanan data berukuran cukup besar. Istilah lain dari *data mining* diantaranya *knowledge discovery (mining) in databases* (KDD), *knowledge extraction, data/pattern analysis, data archeology, data dredging, information harvesting, dan business intelligence* [13]. *Data mining* adalah analisa terhadap data untuk mendapatkan hubungan yang jelas dan menyimpulkan bagian yang belum diketahui sebelumnya [14]. Ada beberapa tahapan dalam *data mining*. Fase pertama dimulai dari sumber data dan diakhir menghasilkan informasi, adapun tahapan dalam *data mining* dapat dilihat pada Gambar 2 [15].

p-ISSN: 2252-3006
e-ISSN: 2685-2411



Gambar 2. Tahapan *Data Mining*

Gambar 2 menunjukan bahwa tahapan *data mining* terdiri atas lima tahapan diantaranya seleksi data, *preprocessing*, transformasi, *data mining*, dan interpretasi / evaluasi [16].

3.3 K-Means Clustering

Metode K-Means *Clustering* merupakan salah satu metode pengelompokkan data non hirarki yang dapat mempartisi data ke dalam *cluster* [17]. Metode ini mempartisi data ke dalam *cluster* sehingga data dengan karakteristik yang sama dikelompokkan ke dalam satu *cluster* yang sama dan data dengan karateristik yang berbeda dikelompokkan ke dalam *cluster* lain. Tahapan awal dari metode ini adalah menentukan jumlah pusat *cluster* dari dataset. Penentuan titik *centroid cluster* awal dilakukan secara acak, kemudian secara iteratif operasi dilakukan hingga tidak ada data yang berpindah *cluster* atau tidak ada perubahan nilai *centroid*. Algoritma K-Means menggunakan rumus *Euclidean* untuk menghitung jarak data pada *data log* [18].

3.4 Silhouette Ceofficient

Silhouette Coefficient digunakan untuk mengetahui kualitas serta kekuatan dari setiap *cluster*, seberapa cocok objek yang ditempatkan dalam suatu *cluster*. Metode *Silhouette Coefficient* digunakan untuk mengetahui kualitas dari hasil *clustering* metode K-Means. Metode ini merupakan gabungan dari metode *cohesion* dan *separation* [19]. Tahapan perhitungan dari metode *Silhouette Coefficient* adalah sebagai berikut.

1. Hitung rata-rata jarak dari suatu data misalkan *i* dengan semua data lain yang berada dalam satu *cluster* (*ai*).

$$a(i) = \frac{1}{|A| - 1} \sum_{j} \in_{A, j \neq i} d(i, j) \quad (5)$$

Diketahui *j* merupakan objek lain dalam satu *cluster A* dan *d(i,j*) merupakan jarak antara objek *i* dengan *j*.

2. Hitung rata-rata jarak dari objek *i* tersebut dengan semua objek pada *cluster* lain dan diambil nilai paling terkecil.

$$d(i, C) = \frac{1}{|A|} \sum j \in C \, d(i, j) \quad (6)$$

Diketahui *d(j,C)* merupakan jarak rata-rata data *i* dengan semua objek pada *cluster* lain dimana *A* tidak sama dengan *C*.

3. Perhitungan selanjutnya yaitu menghitung *Silhouette Coefficient* dengan rumus sebagai berikut.

$$s(i) = (b(i) - a(i)) / \max(a(i), (b(i)) \quad (7)$$

Diketahui bahwa rumus *s(i)* merupakan semua rata-rata pada semua kumpulan objek. *a(i)* merupakan rata-rata objek *i* dengan semua objek lain yang berada dalam satu *cluster*. *b(i)* adalah rata-rata jarak objek *i* dengan semua objek pada *cluster* lain dan diambil nilai paling kecil.

p-ISSN: 2252-3006
e-ISSN: 2685-2411

4. Hasil dan Pembahasan

Bagian hasil dan pembahasan membahas mengenai capaian dari penelitian ini. Hasil dibagi menjadi hasil frekuensi serangan, hasil *clustering*, perhitungan nilai risiko, dan analisa dilakukan terhadap visualisasi kategori risiko serangan.

4.1 Frekuensi Serangan

Data yang digunakan dalam *clustering* terdiri dari 308.781 data serangan yang terdeteksi oleh *honeypot*. Data serangan yang akan diolah sudah dalam bentuk *csv file*. Adapun parameter *cluster* berjumlah 3 *cluster*. Sebelum dilakukan proses *clustering*, data serangan diolah untuk mencari frekuensi dari masing-masing serangan. Frekuensi serangan yang paling tertinggi dengan tujuan protokol "*mssqld*" dimana frekuensi serangan mencapai 306.743 yang dapat dilihat pada Tabel 3.

Tabel 3. Frekuensi Serangan

Protokol	Frekuensi
mssqld	306743
telnet	1187
httpd	337
SipSession	160
smbd	104
pptpd	103
SipCall	72
dns	51
mysqld	13
ftpd	7
RtpUdpStream	4

Tabel 3 merupakan frekuensi serangan terhadap jenis layanan. Terlihat bahwa jenis layanan yang paling banyak terkena serangan adalah jenis layanan *mssqld* dengan jumlah serangan hingga 306.743 kali. Sedangkan serangan yang paling sedikit terkena serangan ialah jenis layanan *RtpUdpStream* dengan jumlah serangan hanya 4 kali.

4.2 Hasil Clustering

Hasil *clustering* metode K-Means menghasilkan pengelompokan data kedalam tiga *cluster* dimana setiap pusat *centroid* digunakan untuk memperoleh kategori nilai risiko serangan. Hasil dari proses *clustering* menggunakan metode K-Means berdasarkan *dataset* dapat dilihat pada Tabel 4.

Tabel 4. Hasil Clustering

Cluster	**Jumlah Data**
Cluster 1	306743
Cluster 2	1685
Cluster 3	353

p-ISSN: 2252-3006
e-ISSN: 2685-2411

Tabel 4 menunjukkan jumlah data dari setiap *cluster* berdasarkan hasil *clustering* yang dilakukan. Terlihat bahwa *cluster* 1 memiliki jumlah data yang paling banyak diantara *cluster* lainnya. Adapun jumlah data yang termasuk dalam Cluster 1 berjumlah 306.743 data, jumlah data pada Cluster 2 adalah 1.685 data, dan Cluster 3 memiliki jumlah data yang paling sedikit dengan 353 data. Titik pusat *centroid* akhir dari hasil *clustering* dapat dilihat pada Tabel 5.

Tabel 5. Pusat *Centroid* Akhir

Cluster	Prioritas	Port	Frekuensi
Cluster 1	2	4	5
Cluster 2	3.0652819	4	1
Cluster 3	3.25212465	1.99150142	1

Tabel 5 menunjukkan pusat *centroid* akhir dari setiap *cluster* pada masing-masing parameter. Nilai dari pusat *centroid* akhir didapat saat proses iterasi berhenti berjalan. *Cluster* 1 pada parameter prioritas memiliki nilai pusat *centroid* akhir yaitu 2, parameter *port* memiliki nilai *centroid* akhir yaitu 4, dan parameter frekuensi memiliki nilai *centroid* akhir yaitu 5. *Cluster* 2 pada parameter prioritas memiliki nilai pusat *centroid* akhir yaitu 3.0652819, parameter *port* memiliki nilai *centroid* akhir yaitu 4, dan parameter frekuensi memiliki nilai *centroid* akhir yaitu 1. *Cluster* 3 pada parameter prioritas memiliki nilai pusat *centroid* akhir yaitu 3.25212465, parameter *port* memiliki nilai *centroid* akhir yaitu 1.99150142, dan parameter frekuensi memiliki nilai *centroid* akhir yaitu 1.

4.3 Perhitungan Nilai Risiko

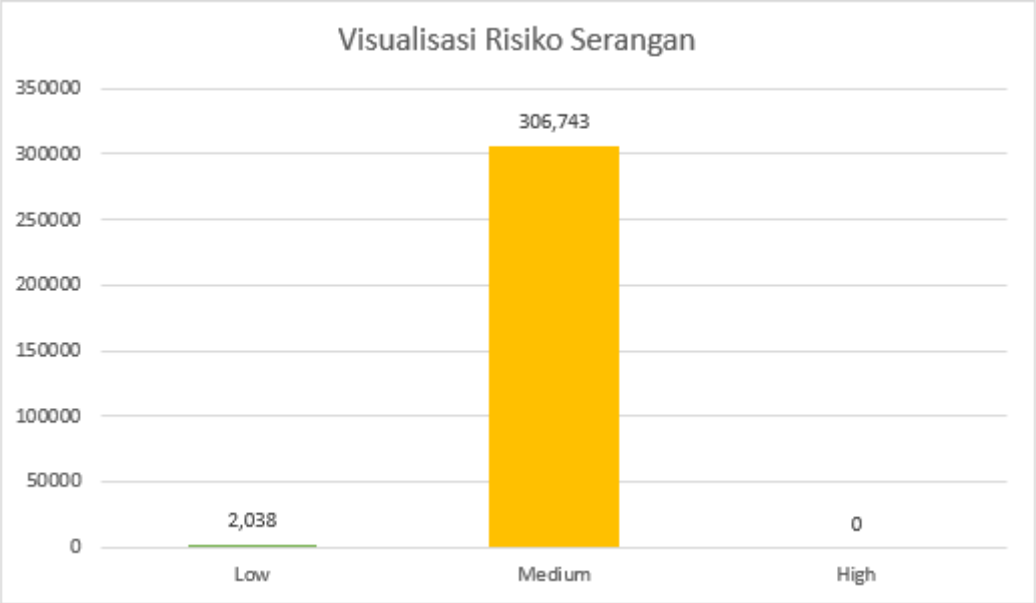
Pusat *centroid* akhir dari setiap *cluster* digunakan untuk perhitungan nilai risiko dari setiap serangan dengan menggunakan Persamaan 4. Hasil perhitungan nilai risiko dan kategori dari setiap *cluster* dapat dilihat pada Tabel 6.

Tabel 6. Hasil Perhitungan Nilai Risiko

Cluster	*P*	*D*	*F*	RA	Kategori
Cluster 1	2	4	5	5.0	Medium
Cluster 2	3.0652819	4	1	1.5	Low
Cluster 3	3.25212465	1.99150142	1	0.8	Low

Tabel 6 merupakan hasil nilai risiko yang diperoleh berdasarkan pusat *centroid* akhir dari ketiga parameter. Adapun hasil dari metode K-Means *Clustering* menghasilkan dua kategori diantaranya *Cluster* 2 dan *Cluster* 3 masuk kedalam kategori serangan *Low* dan *Cluster* 1 masuk kedalam kategori serangan *Medium*. Terlihat bahwa *Cluster* 2 dan *Cluster* 3 masuk dalam kategori yang sama dikarenakan nilai (*RA*) dari masing-masing *cluster* masuk dalam kategori risiko serangan *Low* dimana nilai (*RA*) dari *Cluster* 2 adalah 1.5 dan *Cluster* 3 adalah 0.8, sedangkan *Cluster* 1 masuk dalam kategori *High* dikarenakan nilai (*RA*) berjumlah 5.0. Visualisasi risiko serangan dapat dilihat pada Gambar 3.

p-ISSN: 2252-3006
e-ISSN: 2685-2411



Gambar 3. Visualisasi Risiko Serangan

Gambar 3 merupakan visualisasi dari setiap kategori risiko serangan. Data serangan dengan jumlah terbesar berada pada kategori risiko serangan *Medium* dengan jumlah hingga 306.743 data, sedangkan data serangan yang masuk pada kategori risiko serangan *Low* berjumlah 2.038 data. Tidak ada data serangan yang masuk kedalam kategori risiko serangan *High* dikarenakan perhitungan nilai risiko pada hasil *clustering* tidak ada yang mendekati dengan kategori tersebut.

Hasil *clustering* menggunakan 308.781 data serangan yang dievaluasi dengan menggunakan metode *silhouette score* yang merupakan nilai rata-rata dari perhitungan metode *silhouette coefficient*. Adapun nilai *silhouette score* dari hasil *clustering* metode K-Means adalah 0.999.

5. Kesimpulan

Pengelompokan data serangan dengan memanfaatkan metode K-Means *Clustering* merupakan salah satu cara untuk memahami data serangan yang tersimpan di sebuah *data log*, dikarenakan kemudahan dari implementasi dan sesuai dengan karakteristik data dari *data log honeypot*. Dapat dilihat protokol mana saja yang sering diserang serta kategori risiko serangan yang terdeteksi pada *data log honeypot*. Metode K-Means biasanya digunakan dalam beberapa penelitian lain dalam pengelompokan *data log* dan memiliki kualitas hasil yang baik serta relatif cepat untuk dijalankan. Hasil *clustering* yang telah dilakukan dengan jumlah 308.781 data dapat menentukan kategori serangan dengan baik dimana dari tiga *cluster*, *Cluster* 2 dan *Cluster* 3 berada pada kategori risiko serangan *Low* dengan jumlah 2.038 data, sedangkan *Cluster* 1 masuk pada kategori risiko serangan *Medium* dengan jumlah 306.743 data dengan nilai *silhouette score* ialah 0.999 yang menandakan bahwa hasil *clustering* telah sesuai dan data telah cocok dengan *cluster* yang ditempati.

Daftar Pustaka

[1] Made, Ngurah, Nyoman Piarsa, and Arya Sasmita. 2018. “Telegram Bot Integration with Face Recognition as Smart Home Features.” *International Journal of Computer Applications*, Vol.182, No.13: 42–47.

[2] Huang, Cheng, Jiaxuan Han, Xing Zhang, and Jiayong Liu. 2019. “Automatic Identification of Honeypot Server Using Machine Learning Techniques.” *Security and Communication Networks* 2019.

[3] Sulistya, I Made Ari, and Gusti Made Arya Sasmita. 2020. “Network Security Monitoring System on Snort with Bot Telegram as a Notification.” *International Journal of Computer Applications Technology and Research*, Vol.9, No.2: 059–064.

[4] Mokube, Iyatiti, and Michele Adams. “Honeypots : Concepts , Approaches , and Challenges.” *Proceedings of the 45th Annual Southeast Regional Conference*: 321–26.

[5] Fan, Wenjun, Zhihui Du, Max Smith-Creasey, and David Fernandez. 2019. “HoneyDOC: An Efficient Honeypot Architecture Enabling All-Round Design.” *IEEE Journal on Selected Areas in Communications*,. Vol. 37 No.3: 683–97.

[6] Mansoori, Masood, Omar Zakaria, and Abdullah Gani. 2012. “Improving Exposure of Intrusion Deception System through Implementation of Hybrid Honeypot.” *International Arab Journal of Information Technology*, Vol.9, No.5: 436-444.

[7] Fadhilah Dhinur Aini, Imam Riadi dan Rusydi Umar. 2018. “Perancangan Deteksi Anomali Traffic Untuk Investigasi Log Menggunakan Metode K-Means Clusters.” *Prosiding Seminar Nasional Sains dan* Teknologi: 69–74.

[8] Zulfadhilah, Muhammad, Imam Riadi, and Yudi Prayudi. 2016. “Log Classification Using K-Means Clustering for Identify Internet User Behaviors.” *International Journal of Computer Applications*, Vol.154, No.3: 34–39.

[9] NIST. 2012. “NIST Special Publication 800-30 Revision 1 - Guide for Conducting Risk Assessments.” *NIST Special Publication* (September): 95.

[10] El Mostapha, Chakir, Mohamed Moughit, and Youness Idrissi Khamlichi. 2018. “Building an Efficient Alert Management Model for Intrusion Detection Systems.” *Advances in Science, Technology and Engineering Systems*, Vol.3 No.1: 18–24.

[11] Ardianto Setyo Nugroho, Suwanto Raharjo, Joko Triyono. 2013. “Analisis Dan Implementasi Honeypot Menggunakan Honeyd Sebagai Alat Bantu Pengumpulan Informasi Aktivitas Serangan Pada Jaringan.” *Jurnal Jarkom*., Vol. 01, No. 01: 40–48.

[12] Yang, Manzhi, and Qiaoyan Wen. 2017. “Detecting Android Malware by Applying Classification Techniques on Images Patterns.” *2017 2nd IEEE International Conference on Cloud Computing and Big Data Analysis, ICCCBDA 2017*: 344–47.

[13] Bastian, Ade et al. 2018. “Penerapan Algoritma K-Means Clustering Analysis Pada Penyakit Menular Manusia (Studi Kasus Kabupaten Majalengka).” *Jurnal Sistem Informasi (Journal of Information System)* Vol. 14, No. 1: 26–32.

[14] Darmi, Yulia, and Agus Setiawan. 2016. “Penerapan Metode Clustering K-Means Dalam Pengelompokan Penjualan Produk.” *Jurnal Media Infotama*,. Vo.12, No. 2: 148–57.

[15] Gullo, Francesco. 2015. “From Patterns in Data to Knowledge Discovery: What Data Mining Can Do.” *Physics Procedia* 62: 18–22.

[16] Gustientiedina, M. Hasmil Adiya, and Yenny Desnelita. 2019. “Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan Pada RSUD Pekanbaru.” *Jurnal Nasional Teknologi dan Sistem Informasi*, Vol.5, No.1: 17–24.

[17] Yudi Agusta. 2007. “K-Means – Penerapan, Permasalahan Dan Metode Terkait.” *Jurnal Sistem dan Informatika*, Vol.3: 47–60.

[18] Agung, Anak et al. 2019. “Identifikasi Sel Human African Trypanosomiasis Pada Sel Darah Dengan Menggunakan K-Means Clustering.” *J. Ilm. Merpati (Menara Penelit. Akad. Teknol. Informasi)*; Vol. 7, No. 3: 170–81.

[19] Shoolihah, Al-Mar’atush, M. Tanzil Furqon, and Agus Wahyu Widodo. 2017. “Implementasi Metode Improved K-Means Untuk Mengelompokkan Titik Panas Bumi.” *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya*, Vol.1, No.11: 1270–76.