JINTEKS (Jurnal Informatika Teknologi dan Sains) ISSN 2686-3359 (Online)
Vol. 5 No. 1, Februari 2023, hlm. 49 – 53

ANALISA CLUSTER DENGAN K-MEAN CLUSTERING UNTUK PENGELOMPOKAN DATA CYBERCRIME

Wulan Permata Sari[1*], Tata Sutabri[2]
[1,2]Magister Teknik Informatika, Universitas Bina Darma, Palembang, Sumatera Selatan
email: Wulanpermataabd@gmail.com[*]

Abstrak: Tujuan penelitian ini untuk melakukan cluster atau pengelompokan terhadap dataset *cybercrime*. Diketahui potensi kejahatan terkait data sangatlah mungkin untuk terjadi. Beberapa negara telah sejak lama memiliki perhatian yang lebih untuk keamanan data yang ada didalam dunia maya. Pada penelitian ini penulis ingin melakukan pengelompokan data atau clustering terhadap data Cybercrime yang didapat dari dataset kaggle. Untuk itu perlu dilakukan pengelompokan jenis kejahatan siber atau *cybecrime* dnegan menggunakan metode *k-mean clustering* yang dimana metode tersebut adalah suatu algoritma pengklasteran yang cukup sederhana yang mempartisi database kedalam beberapa clasteran k. Mempartisi data yang ada ke dalam bentuk dua atau lebih kelompok. Guna untuk mengetahui hasil cluster dari metode *K-Means Clustering* peneliti menggunakan alat bantu rapidminer, yang dimana didalam alat bantu tersebut sudah ada operator data K-Means Clustering, sehingga dengan tools tersebut akan didapat hasil dari pola pengelompokan data dari dataset *cybercrime* yang telah dikumpulkan oleh penulis. Hasil penelitian ini adalah sebuah pengelompokan data terhadap dataset *cybercrime* yang dibagi kedalam 3 kelompok atau *cluster* yang dimana menghasilkan hasil uji dengan K-Mean Clustering didapati bahwa pola K yang digunakan dari 3 cluster memiliki cluster 1 sebagai cluster yang paling dominan dengan 20 *record data*.

Kata Kunci : Analisa, Kluster, Kejahatan Siber, K-Mean

Abstract: *The purpose of this research is to cluster or group cybercrime datasets. It is known that the potential for data-related crimes is very likely to occur. Several countries have long paid more attention to data security in cyberspace. In this study, the authors wanted to group or cluster data on Cybercrime data obtained from the Kaggle dataset. For this reason, it is necessary to classify types of cybercrime using the k-mean clustering method, which is a fairly simple clustering algorithm that partitions the database into several k clusters. Partition existing data into two or more groups. In order to find out the cluster results from the K-Means Clustering method, researchers use the rapidminer tool, in which the K-Means Clustering data operator already exists in the tool, so that with these tools you will get results from data grouping patterns from cybercrime datasets that have been collected by writer. The results of this study are a grouping of data on cybercrime datasets which are divided into 3 groups or clusters which produce test results with K-Mean Clustering it is found that the K pattern used from 3 clusters has cluster 1 as the most dominant cluster with 20 data records*.

Keywords **:** Analytics, Clusters, Cybercrime, K-Mean

PENDAHULUAN

Perkembangan teknologi informasi saat ini sudah berkembang dengan pesat[1]. Hal yang merisaukan dari perkembangan teknologi informasi yang senantiasa berubah serta cepatnya dari perkembangan software, keamanan merupakan suatu isu yang sangat krusial dan setiap orang mempertaruhkan waktu dan biaya untuk melindungi data privasi di internet[2]. Pada era globalisasi seperti sekarang ini,perkembangan teknologi sangatlah pesat,Teknologi Informasi dan komunikasi (TIK) telah menjadi bagian hidup manusia. Keberadaan TIK membuat hidup menjadi lebih mudah dan menyenangkan. Tetapi TIK juga bisa digunakan untuk tindak kejahatan. Cybercrime adalah suatu tindak kriminal yang dilakukan dengan menggunakan Teknologi komputer sebagai alat kejahatan utama[3].

Potensi kejahatan terkait data sangatlah mungkin untuk terjadi. Beberapa negara telah sejak lama memiliki perhatian yang lebih untuk keamanan data yang ada didalam dunia maya. Implementasi dari perhatian tersebut tertuang dalam regulasi-regulasi nasional terkait teknologi informasi. Indonesia menuangkan segala hak dan kewajiban terkait hukum siber didalam UndangUndang Nomor 19 tahun 2016 tentang Informasi dan Transaksi Elektronik yang disingkat dengan UU ITE[4].

Pada penelitian ini penulis ingin melakukan pengelompokan data atau *clustering* terhadap data *Cybercrime* yang didapat dari dataset *kaggle*. Clustering adalah proses pengelompokan satu set data objek menjadi beberapa kelompok atau klaster sehingga objek dalam sebuah klaster memiliki kemiripan yang tinggi satu sama lain, tetapi sangat berbeda dengan objek dalam kelompok lainnya. Salah satu algoritma yang sering digunakan adalah algoritma k-means[5].

Metode yang digunakan pada penelitian ini adalah metode K-Mean merupakan metode Analisa kelompok yang diarahkan pada pemartisian. N Obyek pengamatan kedalam K Kelomok atau disebut sebagai cluster, dimana setiap obyek pengamatan memiliki sebuah kelompok dengan rata-rata atau mean. K-mean merupakan salah satu metode pengelompokkan data sekatan (nonhierarki) yang berusaha mempartisi data yang ada kedalam bentuk

 ISSN 2686-3359 (Online)

dua atau lebih kelompok. Metode ini mempartisi data ke dalam kelompok sehingga data yang berkarakter berbeda di kelompokkan ke dalam kelompok yang lain[6].

Guna untuk mengetahui hasil cluster dari metode *K-Means Clustering* peneliti menggunakan alat bantu rapidminer, yang dimana didalam alat bantu tersebut sudah ada operator data K-Means Clustering, sehingga dengan tools tersebut akan didapat hasil dari pola pengelompokan data dari dataset *cybercrime* yang telah dikumpulkan oleh penulis.

Salah satu penelitian terkait yang menjadi refrensi penelitian yang sedang dijalankan adalah penelitian Penerapan Data Mining Pengelompokan Kejahatan Elektronik Sesuai UU ITE dengan Menggunakan Metode Clustering, yang dimana dari keseluruhannya dapat di simpulkan dari hasil jenis kriteria usia, pelanggaran dan juga pasal yang dikenakan bahwa pelanggaran yang paling dominan terjadi yaitu Penyebaran Informasi Hoax dengan nilai 2,75 5,87 6,25 kemudian disusul dengan jenis kejahatan Pemerasan/Pengancaman dengan jumlah 3,57 3,42 3,14 serta pelanggaran yang minimum terjadi adalaj Konten pornografi yang bernilaikan sebesar 3,4 1 1 [7].

Kemudian penelitian yang dijalankan Dhika (2021), penelitian ini menerapkan Data Mining dengan menggunakan metode Clustering untuk memetakan strategi pemilihan dearah pemilih pada Calon Legislatif Kota Bengkulu Algoritma yang digunakan yaitu K-Means Clustering, di mana data dikelompokkan berdasarkan karakteristik yang sama akan dimasukkan ke dalam kelompok yang sama dan set data yang dimasukkan ke dalam kelompok tidak tumpang tindih[8].

TINJAUAN PUSTAKA

Analisa

Analisis adalah sebuah kegiatan untukmencari pola ataupun metode berasumsi yang berhubungan dengan pengetesan dengan cara analitis kepada suatu buat memastikan bagian, jalinan antar bagian, dan ikatan dengan totalitas. Analisa merupakan sesuatu upaya buat menguraikan sesuatu permasalahan jadi bagian-bagian(decomposition) alhasil lapisan wujud sesuatuyang dijabarkan itu nampak dengan nyata alhasil dapatdimengerti permasalahannya[9]

Analisis Cluster

Analisis cluster merupakan metode multivariat yang mempunyai tujuan untuk untuk pengelompokkan, dimana suatu kelompok mempunyai ciri yang relatif sama (homogen), sedangkan antar kelompok memiliki ciri yang berbeda. Pada umumnya suatu objek dimasukkan ke dalam suatu klaster atau kelompok sehingga lebih cenderung berhubungan (berkorelasi) dengan objek lainnya di dalam klastermya daripada dengan objek dari klaster lain. Pembentukan klaster didasarkan pada kuat tidaknya hubungan antar objek[10].

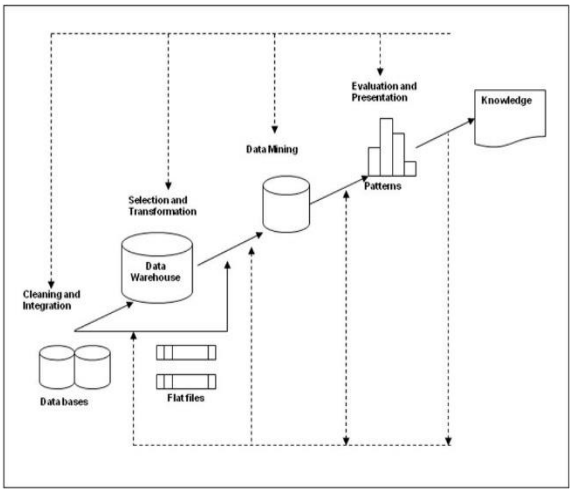
Informasi

Informasi adalah data yang telah diklasifikasikan atau diolah atau diinterpretasikan untuk digunakan dalam proses pengambilan keputusan[11][12].

Data Mining

Data mining, sering juga disebut *Knowledge Discovery In Database (KDD)* adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam data berukuran besar. Keluaran dari data mining ini bisa dipakai untuk memperbaiki pengambilan keputusan di masa depan. Saat ini istilah pengenalan pola (pattern recognition) jarang digunakan karena ia termasuk bagian dari data mining[13].

Knowledge Discovery in Database adalah proses untuk menggali dan menganalisis sejumlah data dan mengestrak informasi dan pengetahuian yang berguna. Hasil pengetahuan yang diperoleh dalam proses tersebut dapat digunakan sebagai basis pengetahuan untuk keperluan pengambilan keputusan". Proses dalam KDD adalah proses yang digambarkan pada dan terdiri dari rangkaian proses sebagai berikut[14]:



Gambar 1. Tahapan KDD

Data mining merupakan salah satu langkah dari proses Knowledge Discovery from Data atau lebih dikenal dengan singkatan KDD". Berikut langkah langkah dari KDD [15]: Tahap-tahap data mining ada 6 yaitu:

1. Pembersihan data (data cleaning)

   Pembersihan data merupakan proses menghilangkan noise dan data yang tidak konsisten atau data tidak relevan. Pada umumnya data yang diperoleh, baik dari database suatu perusahaan maupun hasil

 ISSN 2686-3359 (Online)

eksperimen, memiliki isian-isian yang tidak sempurna seperti data yang hilang, data yang tidak valid atau juga hanya sekedar salah ketik. Selain itu, ada juga atribut-atribut data yang tidak relevan dengan hipotesa data mining yang dimiliki. Data-data yang tidak relevan itu juga lebih baik dibuang. Pembersihan data juga akan mempengaruhi performasi dari teknik data mining karena data yang ditangani akan berkurang jumlah dan kompleksitasnya.

2. Integrasi data (data integration)

   Integrasi data merupakan penggabungan data dari berbagai database ke dalam satu database baru. Tidak jarang data yang diperlukan untuk data mining tidak hanya berasal dari satu database tetapi juga berasal dari beberapa database atau file teks. Integrasi data dilakukan pada atribut-aribut yang mengidentifikasikan entitas-entitas yang unik seperti atribut nama, jenis produk, nomor pelanggan dan lainnya. Integrasi data perlu dilakukan secara cermat karena kesalahan pada integrasi data bisa menghasilkan hasil yang menyimpang dan bahkan menyesatkan pengambilan aksi nantinya. Sebagai contoh bila integrasi data berdasarkan jenis produk ternyata menggabungkan produk dari kategori yang berbeda maka akan didapatkan korelasi antar produk yang sebenarnya tidak ada. .

3. Seleksi Data (*Data Selection*)

   Data yang ada pada database sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari database. Sebagai contoh, sebuah kasus yang meneliti faktor kecenderungan orang membeli dalam kasus market basket analysis, tidak perlu mengambil nama pelanggan, cukup dengan id pelanggan saja.

4. Transformasi data (*Data Transformation*)

   Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam data mining. Beberapa metode data mining membutuhkan format data yang khusus sebelum bisa diaplikasikan. Sebagai contoh beberapa metode standar seperti analisis asosiasi dan clustering hanya bisa menerima input data kategorikal. Karenanya data berupa angka numerik yang berlanjut perlu dibagi-bagi menjadi beberapa interval. Proses ini sering disebut transformasi data.

5. Proses *mining*,

   Merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data.

6. Evaluasi pola (*Pattern Evaluation*),

   Untuk mengidentifikasi pola-pola menarik kedalam *knowledge based* yang ditemukan. Dalam tahap ini hasil dari teknik data mining berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai apakah hipotesa yang ada memang tercapai. Bila ternyata hasil yang diperoleh tidak sesuai hipotesa ada beberapa alternatif yang dapat diambil seperti menjadikannya umpan balik untuk memperbaiki proses data mining, mencoba metode data mining lain yang lebih sesuai, atau menerima hasil ini sebagai suatu hasil yang di luar dugaan yang mungkin bermanfaat.

7. Presentasi pengetahuan (*Knowledge Presentation*),

   Merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna. Tahap terakhir dari proses data mining adalah bagaimana memformulasikan keputusan atau aksi dari hasil analisis yang didapat. Ada kalanya hal ini harus melibatkan orang-orang yang tidak memahami data mining. Karenanya presentasi hasil data mining dalam bentuk pengetahuan yang bisa dipahami semua orang adalah satu tahapan yang diperlukan dalam proses data mining. Dalam presentasi ini, visualisasi juga bisa membantu mengkomunikasikan hasil data mining

METODE

Algoritma adalah susunan yang logis dan sistematis untuk memecahkan suatu masalah atau untuk mencapai suatu tujuan tertentu[16]. Pada penelitian ini penulis menggunakan metode *k-mean clustering* sebagai metode pemecah masalah yang dimana metode ini merupakan salah satu teknik dari salah satu fungsionalitas data mining, algoritma clustering merupakan algoritma pengelompokkan sejumlah data menjadi kelompok–kelompok data tertentu[17].

K-Means merupakan suatu algoritma pengklasteran yang cukup sederhana yang mempartisi database kedalam beberapa clasteran k. Mempartisi data yang ada ke dalam bentuk dua atau lebih kelompok. Metode ini mempartisi data ke dalam kelompok sehingga data berkarakteristik sama dimasukkan ke dalam satu kelompok yang sama dan data yang berkarakteristik berbeda dikelompokkan kedalam kelompok yang lain. Adapun tujuan pengelompokkan data ini adalah untuk

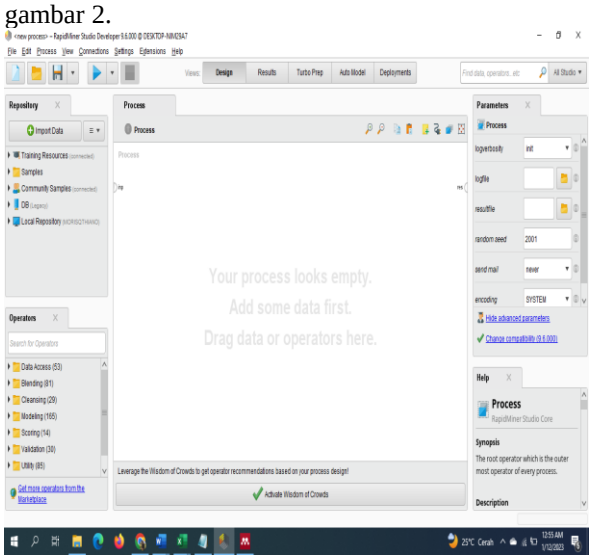
 ISSN 2686-3359 (Online)

meminimalkan fungsi objektif yang diatur dalam proses pengelompokan, yang pada umumnya berusaha meminimalkan variasi di dalam suatu kelompok dan memaksimalkan variasi antar kelompok. Algoritma Kmeans pada dasarnya melakukan 2 proses yakni proses pendeteksian lokasi pusat clusterdan proses pencarian anggota dari tiap-tiap cluster[18].

HASIL DAN PEMBAHASAN

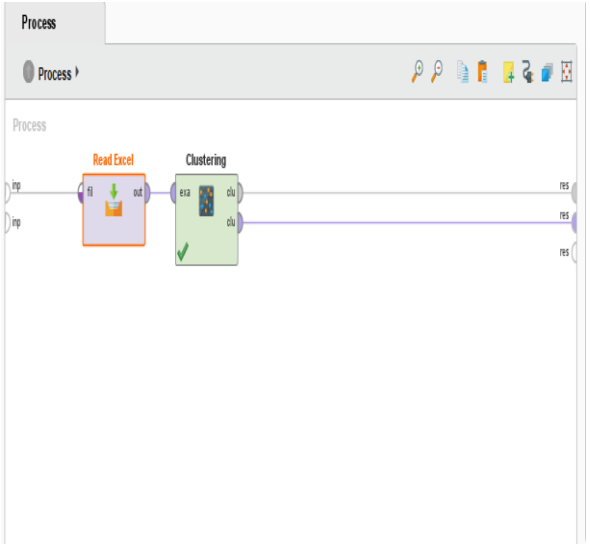
Setelah melakukan analisa terhadap perancangan data mining untuk mengelompokan jenis cluster pada data *cybercrime* yang didapat oleh penulis dari *kaggle* maka proses selanjutnya pengolahan data dengan menggunakan metode metode cluster yaitu metode K-Mean. Analisa ini berakhir dengan proses data mining sesungguhnya, maka hasil yang dicapai adalah pola cluster terhadap kelompok data *cybercrime* dengan metode K-Mean Clustering.

Pada Gambar 2 merupakan antar muka Rapidminer pada antar muka ini terdapat menu, gambar serta tools. Dan untuk memulai pengujian pilih explore pada antar muka kanan atas seperti pada gambar 2.



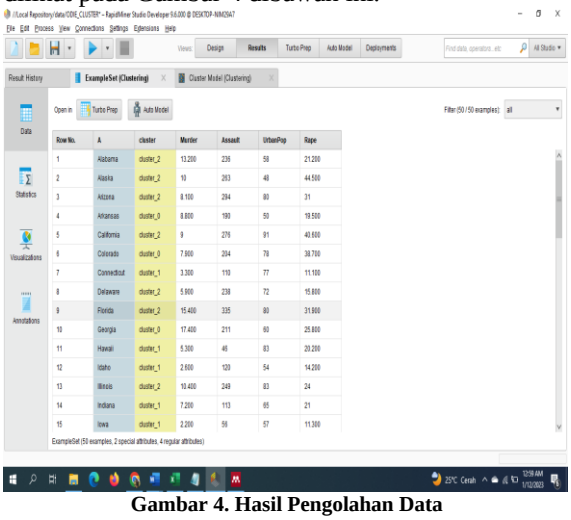
Gambar 2. Menu Rapidminer

Pada langkah berikut ini memilih metode *cluster* yaitu dengan menggunakan *k-mean cluster*, berikut ini merupakan *operator* yang digunakan pada penelitian ini yang dapat dilihat pada Gambar 3.



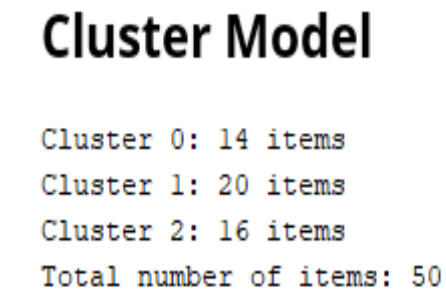
Gambar 3. Operator Menu Rapidminer

Berikut ini merupakan hasil dari dari *model K-Mean Cluster* pada dataset *cybercrime* yang dapat dilihat pada Gambar 4 dibawah ini.



Gambar 4. Hasil Pengolahan Data

Berikut ini merupakan hasil dari dari cluster K-Mean Clustering peneliti menggunakan beberapa kali percobaan dengan menggunakan K=3. Berikut ini hasil dari percobaan tersebut.



Gambar 5. Hasil Cluster

Berdasarkan hasil percobaan diatas dengan menggunakan K=3 atau membagi 3 data cluster maka

 ISSN 2686-3359 (Online)

didapati bahwa pada *cluster 0* memiliki 14 *record* data, sedangkan pada *cluster 1* memiliki 20 *record* data. Terakhir pada *cluster 2* memiliki 16 *record* data. Dari hasil ini dapat disimpulkan bahwa data *cybercrime* didominasi pada kelompok 2 atau *cluster 2.*

KESIMPULAN DAN SARAN

Berdasarkan hasil uji dengan K-Mean Clustering didapati bahwa pola K yang digunakan dari 3 cluster memiliki cluster 1 sebagai cluster yang paling dominan. Sehingga dapat dikatakan data dari *cybercrime* yang digunakan oleh peneliti didominasi oleh data kelompok 2 atau *cluster 1*.

Saran untuk penelitian selanjutnya adalah peneliti mengharapkan untuk dikembangkan kedalam bentuk sistem guna untuk lebih mudah dalam pengelompokan data *cybercrime.* Serta menggunakan dataset yang lebih besar serta bervariatif. Kekurangan peneltiian ini adalah peneliti hanya menggunakan satu metode pemecah masalah sehingga tidak ada perbandingan metode terhadap dataset yang digunakan.

DAFTAR PUSTAKA

[1] T. Sutabri, T. Sugiharto, R. A. Krisdiawan, and M. A. Azis, "Pengembangan Sistem Informasi Monitoring Progres Proyek Properti Berbasis Website Pada PT Peruri Properti," *J. Teknol. Inform. dan Komput.*, vol. 8, no. 2, pp. 17–29, 2022.

[2] S. Rustam, "Analisa Clustering Phising Dengan K-Means Dalam Meningkatkan Keamanan Komputer," *Ilk. J. Ilm.*, vol. 10, no. 2, pp. 175–181, 2018, doi: 10.33096/ilkom.v10i2.309.175-181.

[3] D. Daryono and B. Sugiantoro, "Pengembangan Framework Pelaporan Cyber Crime," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 1, no. 3, pp. 133–147, 2017, doi: 10.14421/jiska.2017.13-05.

[4] N. P. Suci Meinarni and H. B. Sari, "Analisis Potensi Kejahatan di Dalam Dunia Maya Terkait Data," *Kertha Wicaksana*, vol. 14, no. April 2019, pp. 9–15, 2020, [Online]. Available: https://www.ejournal.warmadewa.ac.id/index.php/kertawicaksana/article/view/1530/1355

[5] W. Astuti, A. Widodo, J. T. Elektro, F. Teknik, and U. N. Semarang, "Pemetaan Tindak Kejahatan Jalanan di Kota Semarang Menggunakan Algoritma K-Means Clustering," *J. Tek. Elektro*, vol. 8, no. 1, pp. 5–7, 2016.

[6] A. K. Nalendra, M. Mujiono, A. Rafika, and A. W. Sasama, "IMPLEMENTASI ALGORITMA K-MEAN DALAM PENGELOMPOKAN DATA KECELAKAAN (STUDI KASUS KABUPATEN KEDIRI) Adimas," *Vocat. Educ. Technol. J.*, vol. 1, no. 2, pp. 21–27, 2020, [Online]. Available: http://ojs.aknacehbarat.ac.id/index.php/vocatech/index

[7] M. Simanjuntak and Dkk, "Penerapan Data Mining Pengelompokan Kejahatan Elektronik Sesuai UU ITE dengan Menggunakan Metode Clustering," *J. Mahajana Inf.*, vol. 3, no. 2, p. 3, 2018.

[8] D. Alfatah, "Application of the K-Means Clustering Algorithm in Mapping the Regional Voter Strategy for the Legislative Candidates for the DPR RI Penerapan Algoritma K-Means Clustering dalam Memetakan Strategi Daerah Pemilih pada Calon Legislatif DPR RI," *J. Kom.*, vol. 1, no. 2, pp. 435–443, 2021.

[9] P. M. Purba, A. C. Amandha, R. H. Purnama, and A. Ikhwan, "Analisis Keamanan Website Prodi Sistem Informasi Uinsu Menggunakan Metode Application Scanning," *J. Inform. Teknol. dan Sains*, vol. 4, no. 4, pp. 325–329, 2022.

[10] M. Y. Matdoan, "Penerapan Analisis Cluster Dengan Metode Hierarki Untuk Klasifikasi Kabupaten/Kota Di Provinsi Maluku Berdasarkan Indikator Indeks Pembangunan Manusia," *Statmat J. Stat. Dan Mat.*, vol. 2, no. 2, p. 20, 2020, doi: 10.32493/sm.v2i2.4740.

[11] T. Sutabri, *Analisis Sitem Informasi*, vol. 53, no. 9. 2014.

[12] T. Sutabri, *Konsep Sistem Informasi*. Yogyakarta: Andi, 2012.

[13] D. Novianti, "Implementasi Algoritma Naïve Bayes Pada Data Set Hepatitis Menggunakan Rapid Miner," *Paradig. J. Komput. dan Inform. Univ. Bina Sarana Inform.*, vol. 21, no. 2, pp. 143–148, 2019, doi: 10.31294/p.v20i2.

[14] E. T. L. Kusrini, *Algoritma Data Mining*. Yogyakarta: Andi Offset, 2009.

[15] H. Jiawei and K. Micheline, *Data mining: concepts and techniques second edition*. 2006.

[16] Y. B. Widodo, S. A. Anggraeini, and T. Sutabri, "Perancangan Sistem Pakar Diagnosis Penyakit Diabetes Berbasis Web Menggunakan Algoritma Naive Bayes," vol. 7, no. 1, pp. 112–123, 2021.

[17] S. P. Tamba, F. T. Kesuma, and Feryanto, "Penerapan Data Mining Untuk Menentukan Penjualan Sparepart Toyota Dengan Metode K-Means Clustering," *J. Sist. Inf. Ilmu Komput. Prima (JUSIKOM PRIMA)*, vol. 2, no. 2, pp. 67–72, 2019.

[18] K. F. Irnanda, A. P. Windarto, I. S. Damanik, and I. Gunawan, "Penerapan K-Means pada Proporsi Individu dengan Keterampilan (Teknologi Informasi dan Komunikasi) TIK Menurut Wilayah," no. c, pp. 452–456, 2019.