# *RING* UNTUK PROSES *CYBER PROFILING* : STUDI KASUS PADA DUNIA PENDIDIKAN TINGGI DI INDONESIA

**Muhammad Zulfadhilah[1)2)], Yudi Prayudi[2)], dan Imam Riadi[3)]**
[1)] Program Studi Teknik Informatika Politeknik Hasnur Banjarmasin
[2)] Universitas Islam Indonesia Yogyakarta
[3)]Universitas Ahmad Dahlan
13917220@student.uii.ac.id[1)] prayudi@uii.ac.id[2)]

## ABSTRAK

Aktifitas pengguna internet semakin meningkat dari tahun ke tahun.Hal ini telah berdampak pada perilaku pengguna itu sendiri.Penilaian terhadap perilaku pengguna seringkali hanya didasarkan pada interaksi di internet tanpa mengetahui apa saja aktifitas lainnnya. Penggunaan log aktifitas dapat dijadikan sebagai cara lain untuk mengetahui perilaku pengguna. Log aktifitas internet adalah salah satu tipe big data sehingga penggunaan data mining dengan teknik K-Means dapat dijadikan sebagai solusi untuk melakukan analisis terhadap perilaku pengguna. Pada penelitian ini telah dilakukan proses klasterisasi menggunakan algoritma K-Means yang terbagi menjadi tiga klaster yaitu tinggi, sedang, dan rendah. Hasil yang didapat dari institusi pendidikan menunjukkan bahwa setiap klaster menghasilkan website yang sering dikunjungi dengan urutan : wesite pencarian, social media, berita, dan informasi. Penelitian ini juga menunjukkan bahwa Cyber profiling yang telah dilakukan sangat dipengaruhi oleh faktor lingkungan dan aktifitas sehari-hari.

*Kata Kunci : Clustering, K-Means, Log, Network, Profiling*

## PENDAHULUAN

Semakin banyaknya aplikasi, perangkat keras (device), dan koneksi internet telah berdampak pada perilaku penggunannya.Dalam hal ini hasil riset APJI menunjukkan bahwa pada tahun 2014 urutan aktifitas pengguna internet di Indonesia adalah : pengguna jejaring sosial (*social media*), pencarian informasi, chatting (*messaging*), pencarian berita, video, *email* sebagai urutan popularitas aktifitas pengguna internet. Data tersebut juga menunjukkan bahwa pencarian berita dan penggunaan *email* bukan merupakan aktifitas yang popular (APJII, 2015).

Secara umum penelitian *cyber profiling* adalah melakukan eksplorasi data dengan tujuan untuk mengetahui apa saja aktifitas pengguna pada saat melakukan akses internet.

Salah satu metode yang dapat digunakan untuk mendukung proses profiling adalah algoritma *K-Means*. Melalui algoritma ini maka data dapat dikelompokkan berdasarkan tingkat kunjungannya. Pengelompokkan ini bertujuan untuk melihat website apa saja yang sering diakses oleh pengguna.

Data pengguna akses internet pada suatu institusi dapat dikategorikan sebagai tipe data besar, sehingga proses analisisnya dapat dilakukan dengan data mining. Dalam hal ini algoritma *cluster* sebagai salah satu teknik data mining dapat digunakan untuk menemukan kelompok (*cluster*) objek yang berguna, dimana penggunaannya tergantung dari tujuan analisa data (Hermawati, 2013). Analisis*clustering* adalahsalah satu

metodeyang palingberguna dalamakuisisi pengetahuan, dan digunakan untukmenemukan*cluster*yang mendasardanpolayang penting untuk didistribusikandari dataitu sendiri (Chunchun, dkk., 2011).

Profilingmerupakan proses mengumpulkan datatentang individudan kelompok yang dapat menghasilkan sesuatu yang menarik, mengejutkan danbermakna, korelasibahwa dengan menggunakan mesinyang memilikikekuatanperhitunganyang baik dapatmendeteksi data-data tersebut, sementara kitasebagai manusiatidak bisa (van den Berg, dkk., 2013). *Cyber profiling* membawa langkah yang baik pada ilmu komputer forensik, berdasarkan dari pengalaman yang telah diraih dalam proses penanganan yang telah dilakukan (Irvine, 2010).

Diantara kelompok pengguna internet yang paling banyak adalah institusi pendidikan. Perilaku pengguna pada institusi pendidikan perlu juga dilakukan *profiling* untuk mengetahui karakteristik pengguna dan akses apa saja yang dilakukan. Untuk wilayah Indonesia belum adanya penelitian terkait dengan ini, karena itu *profiling*akan sangat berguna untuk mengetahui perilaku pengguna internet di institusi pendidikan di Indonesia.

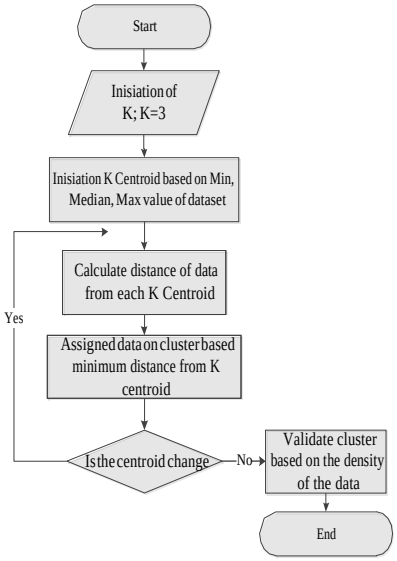
Penggunaan internet di dunia pendidikan seharusnya dimanfaatkan oleh pengguna untuk mendukung proses pendidikan, namun kadang kala didapatkan fakta *user* dilingkungan dunia pendidikan memanfaatkan internet untuk kepentingan di luar dunia pendidikan, bahkan tidak jarang terdapat indikasi user pada institusi pendidikan mengarah pada cyber crime. Untuk itu perlu diketahui lebih lanjut apakah memang penggunaan internet pada dunia pendidikan sudah sesuai dengan lingkup aktifitas di dunia pendidikan.

## METODE PENELITIAN

Untuk mengetahui sejauh mana *cyber profiling* terhadap institusi pendidikan, maka dalam penelitian ini dipilih log data internet dari sebuah institusi pendidikan. Log data tersebut tidak hanya berisikan website apa saja yang diakses oleh pengguna, tetapi juga memuat paket-paket yang diterima dan dikirim melalui lalu lintas jaringan. Data yang didapat berisikan aktifitas lalu lintas jaringan selama lima hari dan menghasilkan data sebanyak 320773 record.

Pada tahap awal penelitian dilakukan pengambilan data.Selanjutnya dilakukan *preprocessing* agar data yang tidak memenuhi kriteria dapat dihilangkan.Pada penelitian ini dari data awal 320773 dihasilkan 1630 record hasil *preprocessing*.Selanjutnya dilakukan mekanisme *clustering* menggunakan Algoritma K-Means yang berjalan pada aplikasi Rapid Miner dan SPSS. Data klaster tersebut kemudian dianalisa untuk melakukan proses profiling terhadap pengguna internet.

Gambar 1 merupakan alur dari penerapan algoritma K-Means dalam proses profling.




Gambar 1.Algoritma K-Means

## HASIL DAN PEMBAHASAN

### *K-Means Clustering*

Pada penerapan algoritma K-Means, hasil yang didapatkan adalah tingkat kunjungan terhadap website. Tingkat kunjungan ini dibagi menjadi tiga kelompok, yaitu rendah, sedang, dan tinggi

Hasil klastering sebagai output dari aplikasi Rapid Miner dan SPSS menunjukkan data klaster yang sama. Berdasarkan hasil *cluster*, terlihat ada 3 *cluster* yang nilainya berbeda-beda, bahkan pada *cluster* kesatu nilainya mencapai 1479 (90,30 %), kedua bernilai 126 (7,70%), dan yang ketiga bernilai 33 (2,00%). Nilai-nilai tersebut merupakan jumlah *website* yang telah terbagi pada masing-masing *cluster*. Hasil klastering yang telah dilakukan menunjukkan proses ini telah berjalan sesuai dengan yang diharapkan pada penelitian.

Inisialisasi pusat *cluster* awal pada proses klastering ini bisa dilihat pada Tabel 1.

Tabel 1. Inisialisasi Pusat Awal Klaster

| Klaster | Jumlah Pengunjung |
|---|---|
| 1 | 1 |
| 2 | 37 |
| 3 | 71 |

Nilai inisialisasi diatas ditentukan berdasarkan dari data yang mempunyai nilai tertinggi, nilai tengah, dan nilai terkecil. Nilai-nilai tersebut menjadi pusat awal *cluster* yang nantinya akan dilanjutkan pada proses *K-Means*.

Perhitungan *centroid* baru akan terus dilakukan (iterasi) sampai dengan ditemukannya iterasi yang mana hasil *centroid*-nya sama dengan hasil *centroid* sebelumnya. Pada penelitian ini terdapat 8 iterasi untuk menentukan hasil *cluster* yang tepat dari 1638 objek yang diteliti. Proses iterasi bisa dilihat pada Tabel 2.

Tabel 2. Proses Clustering

| Iterasi | Perubahan Pusat Klaster | | |
|---|---|---|---|
| | 1 | 2 | 3 |
| 1 | 1,522 | 6,620 | 10,429 |
| 2 | 0,150 | 3,805 | 4,857 |
| 3 | 0,147 | 3,173 | 4,000 |
| 4 | 0.158 | 2,332 | 2,194 |
| 5 | 0,060 | 1,221 | 1,727 |
| 6 | 0,067 | 1,109 | 1,262 |
| 7 | 0,000 | 0,113 | 0,410 |
| 8 | 0,000 | 0,000 | 0,000 |

Adapun hasil akhir dari proses iterasidalammenentukan pusat awal klastering dapat terlihat pada Tabel 3.

Tabel 3. Hasil Akhir Pusat Klaster

| Klaster | Jumlah Pengunjung |
|---|---|
| 1 | 2 |
| 2 | 19 |
| 3 | 46 |

Hasil detail *clustering* akan dijelaskan sebagai berikut :

- *Cluster-1*

Pada hasil klastering yang sudah dilakukan, klaster kesatu memiliki data sebanyak 1467 record.Klaster kesatu memiliki anggota terbanyak, namun klaster ini memiliki nilai yang berada dibawah rata-rata keseluruhan data yang diteliti.Pada klaster kesatu, nilai data berada pada rentang 1-10, karena pada klaster ini data yang ada memiliki tingkat kunjungan yang rendah.Dengan demikian, klaster kesatu dikategorikan pada website yang memiliki tingkat kunjungan paling sedikit dari klaster lainnya.

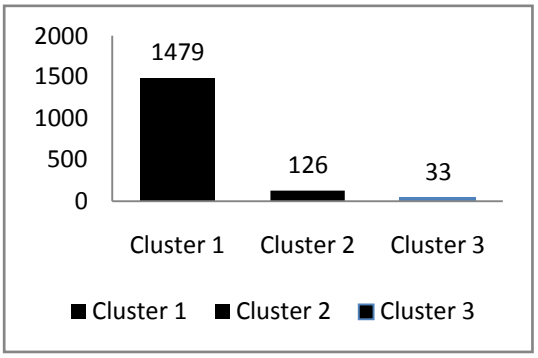
- *Cluster-2*

  Pada klaster kedua, anggota yang masuk pada kalster ini sebanyak 126 record.Nilai dari hasil klaster kedua berada pada rentang 11-31.Nilai ini menunjukkan bahwa anggota dari klaster kedua memiliki tingkat kunjungan yang sedang, karena memiliki nilai sedikit lebih tinggi daripada rata-rata nilai yang dihasilkan oleh klastering.Dengan demikian, klaster kedua dikategorikan sebagai klaster yang memiliki tingkat kunjungan yang sedang.

- *Cluster-3*

  Pada hasil klaster ketiga, anggota yang masuk pada klaster ini sebanyak 33 record.Hasil klaster ketiga ini memiliki jumlah anggota paling sedikit dibandingkan dengan klaster lainnya, namun anggota pada klaster ini memiliki nilai tertinggi data yang telah dihasilkan.Nilai yang ada pada klaster ini berada pada rentang 34-63, hasil ini menunjukka bahwa klaster ketiga memiliki nilai jauh diatas rata-rata.Dengan demikian, klaster ketiga dikategorikan sebagai klaster yang memiliki tingkat kunjungan tertinggi.

Hasil klastering yang telah dilakukan bisa dilihat pada Gambar 2.




Gambar 2.Hasil Klastering

**Analisa Hasil**

Pada penelitian ini algoritma K-Means telah dimplementasikan untuk melakukan klastering sesuai dengan harapan.Pada tahap awal data utama yang didapat berisikan informasi tentang website yang diakses oleh pengguna melalui internet.Selain data informasi website terdapat juga data yang berisikan pembaruan terhadap sistem operasi, pembaruan dari web browser, dan website periklanan yang biasanya tampil sebagai *pop-up*.

Berdasarkan hasil dari K-Means sebagaimana ditunjukkan pada Gambar 1 terlihat bahwa setiap klaster yang didapat memiliki perbedaan jumlah anggota yang signifikan.

Pada klaster kesatu menunjukkan tingkat kunjungan rendah namun memiliki jumlah website yang paling banyak.Data pada klaster kesatu kebanyakan berisi website media periklanan yang muncul bersamaan dengan aktifitas kunjungan pada sebuah website.Sementara itu pada klaster kedua yang memiliki tingkat kunjungan sedang, data anggota klaster menunjukkan adanya situs berita yang berada pada klaster ini.

Pada hasil klaster ketiga yang merupakan kelompok website dengan tingkat kunjungan tertinggi namun memiliki jumlah website yang sedikit.Data pada klaster ini menunjukkan bahwa media sosial

menjadi website dengan tingkat kunjungan yang relative tinggi. Data lain dari klaster tiga menunjukkan bahwa pengguna internet mengakses website pencarian lebih sering terlihat dibandingkan dari website lainnya termasuk website media sosial.

Walaupun pada penelitian (Andayani, 2007)menyebutkan bahwa algoritma K-Means memiliki kekurangan pada inisialisasi pusat awal, namun pada penelitian ini tidak terdapat kekurangan tersebut. Klastering dari dua aplikasi Rapid Miner dan SPSS menghasilkan data yang sama, hal ini menunjukkan bahwa penentuan nilai pusat awal klaster dari kedua aplikasi ini mempunyai nilai awal yang sama.

Berdasarkan hasil klastering yang telah didapat menunjukkan bahwa pengguna internet pada institusi pendidikan sering melakukan akses terhadap website pencarian untuk mendapatkan informasi terkait dengan bidangnya.Selain itu pada penelitian ini juga didapat hasil bahwa website sosial media dan situs video streaming lebih sering diakses dibandingkan dengan website informasi dan berita.Hasil tersebut disebabkan oleh penyampaian informasi pada era digital sekarang lebih cepat memasuki ranah sosial media dibandingkan dengan website informasi lainnya.

Pada penelitian ini, website sosial media termasuk dalam website yang sering diakses oleh pengguna. Hal ini sesuai dengan penelitian yang dilakukan oleh (APJII, 2015; Peña, dkk., 2013)yang menyatakan bahwa sosial media menjadi salah satu aktifitas berinternet yang sering diakses.

Proses *cyber profiling* yang telah dilakukan menunjukkan bahwa pencarian informasi lebih sering diakses oleh pengguna yang berasal dari institusi pendidikan. Hal ini menunjukkan bahwa faktor lingkungan dan aktifitas harian mempengaruhi terhadap apa yang diakses oleh pengguna. Hasil tersebut merujuk pada penelitian (Yu, 2013)yang menyatakan bahwa proses *cyber profiling* dalam memprediksi informasi berdasarkandemografis memiliki tingkat akurasi tinggi.

Berdasarkan data hasil *profiling* terhadap institusi pendidikan menunjukkan bahwa penggunaan internet telah dimanfaatkan untuk mendukung proses pendidikan. Dan pada sumber data yang didapat menunjukkan tidak ada aktifitas user di dunia pendidikan yang mengarah pada *cyber crime.*

Pada penelitian ini sumber data yang diambil masih memiliki kekurangan untuk melakukan *cyber profiling*.

Dalam hal ini seharusnya untuk kelengkapan profiling sebagaimana yang disebutkan oleh (Irvine, 2010)sumber data yang diperlukan untuk seharusnya juga memuat data log aktifitas pada komputer yang telah digunakan, namun data yang didapat hanya berupa data log lalu lintas jaringan.

Pada penelitian ini cyber profiling hanya bisa memberikan informasi tentang aktifitas internet yang dilakukan pengguna, namun untuk ancaman kejahatan dan profiling berdasarkan data komputer yang telah digunakan seperti yang disebutkan oleh (Irvine, 2010) masih belum ada pada penelitian ini.

Hasil dari penelitian ini telah dapat memenuhi pengertian dari *cyber profiling*, karena sudah memberikan informasi tentang pengguna berdasarkan aktifitas saat terhubung dengan internet.Hasil penelitian ini dapat digunakan administrator jaringan untuk keperluan peningkatan kualitas layanan, kebijakan, dan keamanan

seperti yang disebutkan oleh (Bakhshi and Ghita, 2016).

## KESIMPULAN

Dari hasil analisa log dataset dengan menggunakan algoritma *K-Means* untuk proses *cyber profiling*, hasil yang didapat menunjukkan bahwa algoritma initelah dapat mengelompokkan aktifitas pengguna internet berdasarkan pada data kunjungan website. Pengelompokkan ini dibagi menjadi tiga, yaitu kunjungan rendah, sedang, dan tinggi.

Pada penelitian ini, algoritma K-Means digunakan sebagai sebuah algoritma untuk proses *cyber profiling*. Algoritma K-Means yang digunakan sudah sesuai dengan harapan dari penelitian ini, karena memiliki proses algoritma yang sederhana dengan tingkat akurasi yang baik. Namun algoritma K-Means ini memiliki kekurangan, yaitu pada proses pengambilan nilai inisialisasi pusat awal yang acak. Hal ini bisa menyebabkan terjadinya perbedaan pada hasil *cluster.*

Hasil penelitian ini menunjukkan bahwa pengguna internet pada instansi pendidikan lebih banyak mengakses website pencarian untuk mendapatkan informasi.Selain itu hasil yang didapat juga menunjukkan bahwa media sosial memiliki tingkat kunjungan yang tinggi setelah website pencarian.

Berdasarkan hasil penelitian yang menggunakan data dari instansi pendidikan pada tingkat perguruan tinggi menunjukkan bahwa pengguna internet akan lebih cenderung melakukan akses terhadap website yang sesuai dengan bidangnya.

Penelitian ini memiliki keterbatasan pada sumber data untuk proses profiling. Untuk kesempurnaan proses profiling seharusnya data juga memuat data aktifitas komputer yang telah digunakan. Karena itu penelitian selanjutnya diharapkan dapat melakukan *cyber profiling* yang lebih baik dengan sumber data yang lebih lengkap.

## DAFTAR PUSTAKA


APJII. 2015. *Indonesian Internet User Profile 2014*.

Bakhshi, T dan B. Ghita. 2016. *Traffic Profiling : Evaluating Stability in Multi-Device User Environments*.

Chunchun, H., L. Nianxue, Y. Xiaohong, and S. Wenzhong. 2011. Traffic Flow Data Mining and Evaluation Based on Fuzzy Clustering Techniques. 13 (4) : 344–349.

Hermawati, F.A. 2013. *Data Mining*. CV. Andi Offset. Yogyakarta.

Irvine, J. J. 2010. *Digital Forensic Analysis & Cyber Profiling*. 703 : 1–32.

P. Peña, R. Hoyo, J. Vea-murguía, C. González, and S. Mayo. 2013. *Collective Knowledge Ontology User Profiling for Twitter Automatic User Profiling*. pp. 439–444.

S. Andayani. 2007. *Formation of clusters in Knowledge Discovery in Databases by Algorithm K-Means*.

S. Yu. 2013. Behavioral Evidence Analysis on Facebook: a Test of Cyber-Profiling," *Defendologija*. 16 (33) : 19–30.

van den Berg, D. B., P. dr. A. de Vries. dr. S. van der Hof, M. Kakaris, and A. Theocharidis. 2013. Online Identities , Profiling and Cyber Bullying.