



ANALISIS KINERJA PEGAWAI PUSBINDIKLAT PENELITI LIPI BERDASARKAN POLA PEMANFAATAN INTERNET MELALUI PENDEKATAN WEB USAGE MINING

PERFORMANCE ANALYSIS OF PUSBINDIKLAT PENELITI LIPI EMPLOYEE BASED ON INTERNET USAGE USING WEB USAGE MINING APPROACH

Sutrisno Heru Sukoco¹, Imas Sukaesih Sitanggang², Heru Sukoco³

¹²³ Program Studi Magister Ilmu Komputer

Departemen Ilmu Komputer FMIPA IPB, Kampus IPB Baranangsiang, Jl. Raya Pajajaran, Bogor 16127

¹Pusbindiklat Peneliti LIPI, Jl. Raya Bogor KM 46 Cibinong, Bogor 16911

¹sutrisno.kun@gmail.com

Naskah diterima: 1 Agustus 2018; Direvisi : 29 November 2018 ; Disetujui : 12 Desember 2018

Abstrak

Pengukuran kinerja pegawai dalam penggunaan layanan internet dapat dilakukan sebagai bagian dari penilaian kinerja. Pendekatan *web usage mining* melalui pengamatan rekam jejak akses internet yang tersimpan pada *proxy server* merupakan salah satu cara yang dapat diterapkan untuk memahami perilaku pengguna. Penelitian ini bertujuan untuk mendapatkan gambaran perilaku pegawai Pusbindiklat Peneliti LIPI dalam memanfaatkan layanan internet, mengukur level produktivitas pegawai berdasarkan lama waktu akses terhadap situs yang tidak mendukung pekerjaan dan memetakan kategori situs yang diakses apakah mendukung tugas fungsi jabatannya. Penerapan algoritme *clustering K-Means* digunakan untuk memudahkan memahami pola akses pengguna. Data yang digunakan adalah *log proxy server* dan nilai perilaku pegawai Pusbindiklat Peneliti LIPI periode Agustus-Desember 2016. Hasil penelitian menunjukkan pola pemanfaatan internet oleh pegawai Pusbindiklat Peneliti LIPI belum sepenuhnya mendukung tugas fungsi jabatannya. Sekitar 83% pegawai menggunakan internet untuk mengakses situs yang tidak mendukung pekerjaan berada pada level rendah (0-4 jam per minggu). Berdasarkan hasil tersebut dapat disimpulkan bahwa perilaku penggunaan internet yang dilakukan pegawai Pusbindiklat Peneliti LIPI tidak mempengaruhi produktivitas secara signifikan.

Kata kunci: clustering, K-Means, log proxy server, kinerja pegawai, web usage mining

Abstract

Measurement of employee performance in the use of internet services can be conducted as part of an employee's performance target. Web usage mining approach through observation of internet access records stored in the proxy server can be applied in understanding user behavior. This study aims to obtain an overview of employee behavior in utilizing internet services in Pusbindiklat Peneliti LIPI, measure the level of employee productivity based on the length of time access to sites that do not support the work and map the category of sites accessed to the task duty of an employee. K-Means clustering algorithm is used to group user access patterns. The data used are proxy server logs and employee's performance target in Pusbindiklat Peneliti LIPI in a period of August-December 2016. The results show that the patterns of Internet use by employees Pusbindiklat Peneliti LIPI do not fully support the job function. About 83% of employees use the internet to access sites do not support jobs at low-level access (ranging from 0-4 hours per week). Based on these results, it can be concluded that the behavior of internet use by employees of Pusbindiklat Peneliti LIPI does not affect their productivity significantly.

Keywords: clustering, K-Means, log proxy server, performance of employees, web usage mining

PENDAHULUAN

Penggunaan internet di Indonesia berdasarkan data APJII dari tahun ke tahun terus bertumbuh, jumlah pengguna internet pada tahun 2016 sebanyak 132,7 juta bertambah menjadi 143.26 juta di tahun 2017. Berdasarkan hasil survei tahun 2016 jumlah pengguna internet dari lingkungan pekerja atau wiraswasta mencapai 82,2 juta atau sekitar 62%. Gambaran tersebut menunjukkan internet sudah menjadi kebutuhan bagi sebagian besar masyarakat Indonesia. Layanan internet dinilai dapat memberikan dampak positif dalam peningkatan efektifitas dan efisiensi kerja, tetapi pada kenyataannya terkadang ketersediaan layanan internet yang dimiliki oleh instansi pemerintah tidak serta merta memberikan kontribusi yang besar dikarenakan perilaku kerja pegawainya yang belum taat azas dalam pemanfaatannya. Merujuk pada peraturan pemerintah Nomor 46 Tahun 2011 tentang Penilaian Prestasi Kerja Pegawai Negeri Sipil menyebutkan bahwa Prestasi Kerja terdiri atas Sasaran Kinerja Pegawai (SKP) dan Perilaku Kerja. Unsur perilaku kerja memiliki andil yang cukup besar dalam penilaian prestasi kerja pegawai dengan bobot nilai mencapai 40%. Aspek perilaku kerja meliputi orientasi pelayanan, integritas, komitmen, disiplin, kerjasama dan kepemimpinan.

Pengamatan perilaku dalam pemanfaatan internet selama jam kerja merupakan salah satu cara yang dapat dijadikan pertimbangan untuk menilai perilaku kerja pegawai. Lama waktu akses internet dan jenis situs yang sering dikunjungi dapat dijadikan sebagai salah satu tolak ukur penilaian kinerja pegawai (Fathonah & Hartijasti, 2014). Hasil penelitian yang dilakukan oleh (Yusriani & K.

Suprpto, 2016) menunjukkan tingkat akses masyarakat pada *website* pemerintah masih di bawah akses pada media sosial, streaming hal ini sejalan dengan hasil survei yang dilakukan oleh APJII pada tahun 2016 yang menunjukkan jenis konten paling banyak diakses adalah media sosial, hiburan dan berita. Jenis konten media sosial yang paling banyak diakses berdasarkan urutan terbanyak diakses pada tahun tersebut adalah Facebook, diikuti oleh Instagram, dan Twitter. Kondisi tersebut menggugah pimpinan untuk meninjau kembali layanan internet yang dimiliki, apakah pemanfaatan layanan internet lebih banyak digunakan untuk mengakses *website* yang mendukung pekerjaan ataukah tidak mengingat kondisi *bandwidth* jaringan internet pada tahun 2016 alokasi *bandwidth* yang disediakan kondisinya kekurangan, hanya 50-60 Mbps. Pola pemanfaatan internet dapat digali melalui rekam jejak penggunaan internet yang tersimpan pada server dalam bentuk log, salah satunya adalah *proxy server logs* (Zhang, Dai, & Zhou, 2010). Penerapan sistem *proxy server* sendiri merupakan salah satu kebijakan yang ditetapkan oleh pengelola jaringan di lingkungan LIPI.

Salah satu aplikasi *data mining* yang digunakan untuk menemukan pengetahuan yang tersembunyi pada *fail log* adalah *web usage mining* (Pamutha, Chimphee, Kimpan, & Sanguansat, 2012). Terdapat beberapa algoritme yang mendukung penerapan *web usage mining* antara lain *association rule generation*, *sequential pattern generation*, dan *clustering*. Algoritme *clustering* lazim digunakan untuk menggali informasi dari kumpulan data dan mengelompokkannya berdasarkan kemiripan karakteristik (Chitraa & Thanamani, 2012).

Penerapan *clustering* dengan algoritme K-Means diharapkan dapat menggambarkan karakteristik, pola pemanfaatan layanan internet dan produktivitas pegawai (Xu & Liu, 2010). Beberapa penelitian tentang perilaku pengguna internet pernah dilakukan, antara lain oleh (Zubi, Saleh, & Raiani, 2014) tentang *web mining* untuk memahami perilaku pengguna menggunakan *web log dataset*. Penelitian lain yang dilakukan oleh Roiha, (2017) menunjukkan bahwa kategori *blog*, *online shop* dan media sosial merupakan *website* yang sering dikunjungi oleh pengguna internet di lingkungan pemerintah Kota Surabaya.

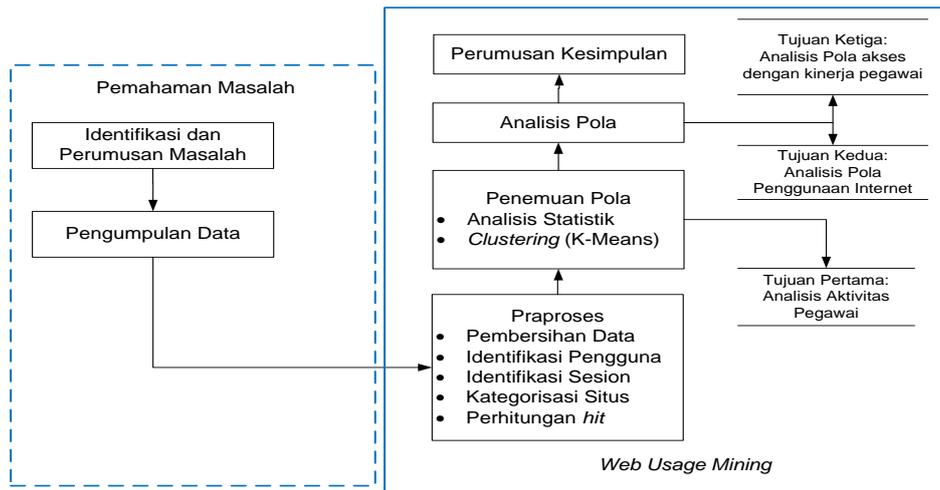
Berdasarkan uraian di atas maka perlu dilakukan penelitian lebih lanjut tentang analisis kinerja pegawai Pusbindiklat Peneliti LIPI berdasarkan pola pemanfaatan internet melalui pendekatan *web usage mining*. Hasil penelitian diharapkan dapat dijadikan sebagai rekomendasi untuk meninjau kembali kebijakan pemanfaatan layanan internet serta sebagai bahan pertimbangan bagi pengambil kebijakan untuk menilai kinerja pegawai terkait pemanfaatan layanan internet yang tersedia apakah sudah sesuai dalam menunjang tugas dan fungsinya.

METODE

Pada penelitian ini menggunakan data primer berupa berkas *log server proxy* LIPI periode bulan Agustus-Desember 2016 dan data Sasaran Kinerja Pegawai (SKP) di lingkungan Pusbindiklat Peneliti LIPI tahun 2016. Sumber data yang digunakan

hanya berasal dari koneksi internet melalui *server proxy*, akses internet di luar *server proxy* tidak termasuk dalam penelitian ini. Data log terkumpul sebanyak 59 *access log file* dengan ukuran yang bervariasi dari mulai 52 MB sampai dengan 8.9 GB, total *access log* yang diperoleh sebesar 146 GB dalam bentuk berkas kompres (.gz).

Penelitian dilaksanakan dalam 2 tahap utama, tahap pertama dimulai dari pemahaman masalah terdiri dari identifikasi dan perumusan masalah untuk mendapatkan data dan informasi mengenai hal yang terkait dengan *web usage mining*, antara lain sumber data *log*, profil pegawai, akun internet (*username*), tahapan dilanjutkan dengan pengumpulan data. Metode yang digunakan untuk pengumpulan data dilakukan dengan cara *copy* rekam jejak akses internet yang tersimpan dalam *server proxy* selama 6 bulan terakhir tahun 2016. Selanjutnya, penggunaan data profil pegawai dan akun internet sudah disesuaikan dengan Peraturan Menteri Komunikasi dan Informatika Nomor 20 Tahun 2016 tentang Perlindungan Data Pribadi dalam Sistem Elektronik, salah satu caranya adalah dengan tidak menyebutkan nama individu secara langsung tetapi hanya menuliskan jabatannya saja guna kepentingan penelitian. Tahap kedua merupakan tahap *web usage mining*, meliputi praproses, *pattern discovery* dan *pattern analysis*, diakhiri perumusan kesimpulan tentang analisis pola pemanfaatan layanan internet dibandingkan dengan kinerja atau produktivitas pegawai. Detail tahapan penelitian ditunjukkan pada Gambar 1.



Gambar 1. Tahapan penelitian

Praproses merupakan tahapan awal dalam *web usage mining* ditujukan untuk mengubah data mentah ke dalam satu set profil pengguna dengan melakukan standardisasi data dan menghilangkan bagian data tertentu yang tidak diperlukan dalam proses *mining* (Dong, 2009). Tahapan praproses meliputi pembersihan data untuk menghilangkan data yang tidak relevan, *noisy* dan entri data yang tidak dibutuhkan. Setelah menghapus baris *log* yang tidak relevan, tahapan selanjutnya adalah memilih *field* yang dibutuhkan. *Field* yang dibutuhkan antara lain *field time* yang merupakan waktu pengaksesan, *username* pengguna, dan URL dari situs yang dikunjungi pengguna. Proses dilanjutkan dengan identifikasi pengguna berdasarkan *username*, tahap berikutnya identifikasi *session* untuk mengidentifikasi aktivitas transaksi berdasarkan selang waktu untuk menghasilkan kualitas temuan pola. Kategorisasi situs ditujukan untuk mendefinisikan situs berdasarkan jenis kategori yang telah ditetapkan. Perhitungan *hit* dilakukan untuk menghitung berapa kali sebuah situs dikunjungi. Tahap kedua *pattern discovery* meliputi analisis statistik dan proses *clustering* dengan algoritme K-Means. Tahap ketiga *pattern*

analysis, dilakukan analisis untuk menilai produktivitas atau kinerja pegawai Pusbindiklat Peneliti LIPI.

HASIL DAN PEMBAHASAN

Praproses Data

Tahap pra-proses mencakup pembersihan data, identifikasi pengguna, *user session identification*, kategorisasi situs, perhitungan *hit* dan perhitungan waktu pemakaian. Tahap praproses cenderung memakan waktu cukup lama terutama diproses kategorisasi, hal ini dikarenakan banyak domain dengan ekstensi .com, co.id, .org, .web.id, dan .net yang memanfaatkan domain tersebut sebagai situs pribadi, jurnal dan lainnya

Pembersihan data

Pembersihan data dilakukan pada baris *record* yang berisi gambar, *script* pendukung, format informasi (iklan), *record* dengan kode status HTTP yang gagal serta metode akses selain GET dan POST (Kerkhofs, Vanhoof, & Pannemans, 2001). Sejumlah record yang tidak sepenuhnya dilakukan oleh pengguna secara langsung seperti *update*

sistem operasi atau antivirus termasuk *record* yang dihapus (Weinreich, Obendorf, & Herder, 2006). Pembersihan baris *log* yang berisi *record* yang dihasilkan oleh mesin pencari otomatis, seperti *Crawler*, *Spider*, dan *Robot* serta *cookies* (Nithya & Sumathi, 2012). Proses tersebut mengurangi jumlah baris *record* sebanyak 29,313,156, jumlah *record* akhir yang tersisa sebanyak 30,406,348 *record*. *Field record* yang diambil dari tahap pembersihan data adalah *IP Address* pengguna, *username* pengguna, URL atau alamat situs yang dikunjungi dan waktu akses (Tanggal, Jam dan Detik).

Identifikasi Pengguna

Identifikasi pengguna didasarkan pada *username*. Baris *record* dengan *username* yang berisi tanda “-“, menggunakan entri data *username* sebelumnya atau melihat kumpulan baris *record* yang sama (Chitraa & Davamani, 2010); (Lüderitz, 2006). Identifikasi pengguna yang dihasilkan sebanyak 9.422 *IP address* yang berbeda dan 3.092 *username*. Pengguna tersebar pada beberapa satuan kerja (satker) yang tersebar di beberapa wilayah di Indonesia, setiap satuan kerja memiliki segmen *IP address* yang berbeda. Identifikasi lokasi pengguna bisa menggunakan lokasi jaringan yang terpasang atau asal satker. Identifikasi jabatan pengguna bisa

diperoleh melalui penelusuran *username* yang didapatkan.

Kategorisasi Situs

Kategorisasi situs ditujukan untuk membantu memudahkan manajemen data (Cadez, Heckerman, Meek, Smyth, & White, 2003). Proses kategorisasi situs dilakukan dengan mengidentifikasi *top level domain* seperti “go.id/ac.id” dan kata kunci tertentu seperti *blogspot*. Rujukan lain untuk kategori situs menggunakan referensi Trustedsource Web Database (Mcafee) dan Palo Alto Networks URL *filtering*. Sebanyak 104 kategori didefinisikan oleh Mcafee, dalam penelitian ini kategori situs dikelompokkan dalam 22 kategori agar klasifikasi lebih fokus. Untuk memudahkan proses, kategori situs didefinisikan dengan memberikan kode kategori seperti ditunjukkan pada Tabel 1.

Perhitungan hit

Perhitungan *hit* dilakukan dengan menghitung berapa kali sebuah situs dikunjungi oleh pengguna. Penjumlahan *hit* dilakukan untuk setiap *username* pegawai. Setelah tahapan praproses selesai diperoleh data berisi jumlah *hit* untuk masing-masing kategori situs berdasarkan lokasi (lokasi jaringan yang terpasang), kedeputian dan satuan kerja pegawai.

Tabel 1. Kategori situs dan jumlah *hitnya*

Kode Kategori	Kategori	Jumlah hit	Kode Kategori	Kategori	Jumlah hit
k1	<i>Blogs/wiki</i>	1,711,502	k12	<i>News</i>	5,454,271
k2	<i>Business</i>	856,119	k13	<i>Online shopping</i>	1,720,820
k3	<i>Education/references</i>	1,256,958	k14	<i>Organization and company</i>	521,244
k4	<i>Entertainment</i>	2,176,330	k15	<i>Pornography</i>	1,240,496
k5	<i>File Sharing</i>	2,974,468	k16	<i>Publication</i>	1,093,571
k6	<i>Finance/Banking</i>	68,167	k17	<i>Real estate</i>	98,982
k7	<i>Gambling</i>	8	k18	<i>Religion</i>	265,629
k8	<i>Games</i>	184,309	k19	<i>Services</i>	2,383,591
k9	<i>Government/military</i>	3,794,102	k20	<i>Social Network/Forum</i>	628,886

k10	<i>Hacking</i>	3,439	k21	<i>Travel</i>	325,717
k11	<i>ICT</i>	726,264	k22	<i>Uncategory</i>	2,920,260

Sumber: data hasil pengolahan

Penemuan Pola

Aktivitas pegawai berdasarkan lokasi kampus

Jumlah *hit* terbanyak berdasarkan lokasi kampus seperti ditunjukkan pada Tabel 2 berada pada Kampus Cibinong, Jakarta, Bandung, dan Serpong. Kondisi tersebut dianggap wajar karena pada kampus tersebut terdiri dari beberapa satker dengan jumlah pegawai yang cukup banyak. Jumlah *hit* yang menarik dan menjadi perhatian adalah Kampus Cibinong dengan jumlah satker dan jumlah pengguna kurang dari jumlah pengguna Kampus

Jakarta tetapi jumlah *hit* terpaut 78,073 hit, jenis kategori yang mendominasi pada Kampus Cibinong adalah *government* (k9) sedangkan pada Kampus Jakarta adalah *news* (k12). Kampus lain yang menarik perhatian adalah Kampus Sukabumi, dengan jumlah satker hanya 1 satker dan pengguna 15 orang tetapi jumlah *hit* mencapai 990,856. Kategori situs yang paling banyak diakses di Kampus Sukabumi adalah kategori *file sharing* (k5) dan *entertainment* (k4).

Tabel 2. Jumlah *hit* berdasarkan lokasi Kampus

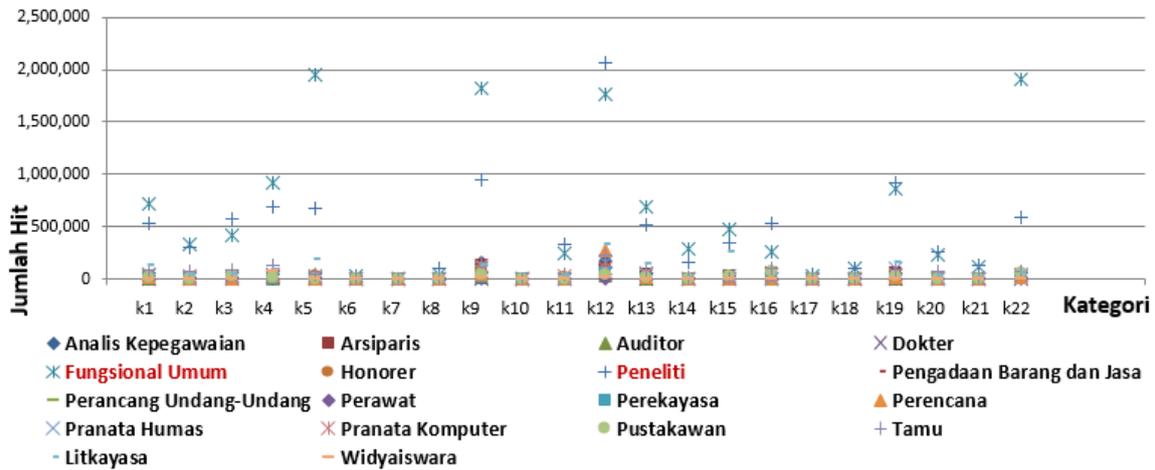
Lokasi	Jumlah Hit	Jumlah Pengguna	Jumlah Satker	Lokasi	Jumlah Hit	Jumlah Pengguna	Jumlah Satker
Cibinong	6,662,258	434	7	Kebumen	438,696	46	1
Jakarta	6,584,185	1027	17	Purwodadi	370,063	85	1
Bandung	5,675,425	632	8	Tual	285,848	13	1
Serpong	5,590,972	614	5	Ambon	265,757	77	1
Bogor	991,234	352	2	Lampung	262,606	55	2
Sukabumi	990,856	15	1	Yogyakarta	215,105	57	1
Cianjur	676,528	73	1	Mataram	194,599	29	1
Subang	652,693	104	1	Bitung	62,693	21	1
Bali	450,816	79	1	Biak	34,799	12	1

Sumber: data hasil pengolahan, lokasi Kampus berdasarkan lokasi jaringan yang terpasang

Aktivitas penggunaan internet berdasarkan kelompok jabatan

Pola pemanfaatan internet berdasarkan kelompok jabatan didominasi oleh pengguna fungsional umum dan fungsional peneliti, seperti ditunjukkan pada Gambar 2. Rataan jumlah *hit* yang dihasilkan oleh pengguna fungsional umum sebanyak 600,821 atau

sekitar 43% sedangkan rataan jumlah *hit* fungsional peneliti sebanyak 32%. Jenis situs yang paling banyak dikunjungi oleh fungsional umum adalah kategori *government* (k9) dan *file sharing* (k5), sedangkan jenis situs yang paling banyak dikunjungi oleh fungsional peneliti adalah kategori *news* (k12) dan *services* (k19).



Gambar 2. Jumlah hit berdasarkan jabatan fungsional di lingkungan LIPI
 Sumber: hasil pengolahan

Tabel 3. Jumlah pegawai berdasarkan kelompok jabatan

No	Jabatan Pengguna	Jumlah	hit	No	Jabatan Pengguna	Jumlah	hit
1	Analisis Kepegawaian (Anpeg)	5	26,306	6	Pranata Komputer (Prakom)	1	22,904
2	Peneliti	1	2,481	7	Widyaiswara (WI)	10	116,573
3	Fungsional Umum (F. Umum)	19	268,952	8	Honorer	1	6,499
4	Perencana	5	8,363	9	Tamu	3	73,713
5	Pranata Humas (Prahum)	7	30,285	Total		52	556,076

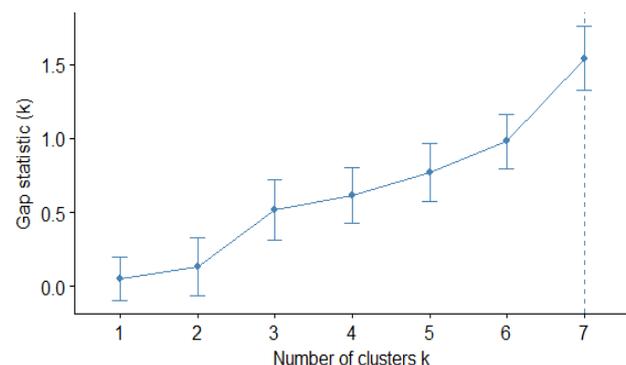
Sumber: hasil pengolahan

Cluster penggunaan internet di Pusbindiklat Peneliti LIPI

Berdasarkan data rekam akses *fail log*, jumlah pengguna yang memanfaatkan akses internet di lingkungan Pusbindiklat Peneliti LIPI periode Agustus sampai Desember 2016 sebanyak 52 pengguna. Jumlah pengguna paling banyak memanfaatkan layanan internet berdasarkan kelompok jabatan adalah fungsional umum yang mencapai 48% dan widyaiswara yang mencapai 21%, seperti ditunjukkan pada Tabel 3.

Karakteristik pengguna internet berdasarkan kelompok jabatan di lingkungan Pusbindiklat dapat dikelompokkan ke dalam dua atau lebih *cluster*. Penentuan jumlah *cluster* optimal dengan kondisi data (jumlah baris yang minim) memungkinkan dilakukan dengan menggunakan metode *gap*

statistik. Kriteria *cluster* optimal ditentukan berdasarkan nilai *gap* statistik (*k*) yang paling tinggi, atau yang pertama kali mengindikasikan kenaikan *gap* minimum jika *gap* selalu naik. Gambar 3 menunjukkan grafik hasil *gap* statistik semakin meningkat hingga mencapai jumlah kluster maksimal. Pilihan yang ideal untuk kondisi tersebut adalah *k*=3 karena nilai tersebut merupakan titik yang pertama kali menunjukkan perbedaan *gap* minimum.



Gambar 3. Nilai k optimal dengan gap statistik

Tabel 4. karakteristik cluster berdasarkan kelompok jabatan

Cluster	Rendah	Sedang	Tinggi	Anggota
1	k10, k18	k1, k2, k3, k5, k6, k8, k9, k11, k12, k13, k14, k15, k16, k17, k19, k20, k21, k22	k4	Widyaiswara
2	k1, k2, k3, k4, k5, k6, k8, k9, k11, k12, k13, k14, k15, k16, k17, k19, k20, k21, k22	k18	k10	Peneliti Perencana Pranata Humas Pranata Komputer Honorar Tamuh
3		k10	k1, k2, k3, k4, k5, k6, k8, k9, k11, k12, k13, k14, k15, k16, k17, k18, k19, k20, k21, k22	Fungsional umum

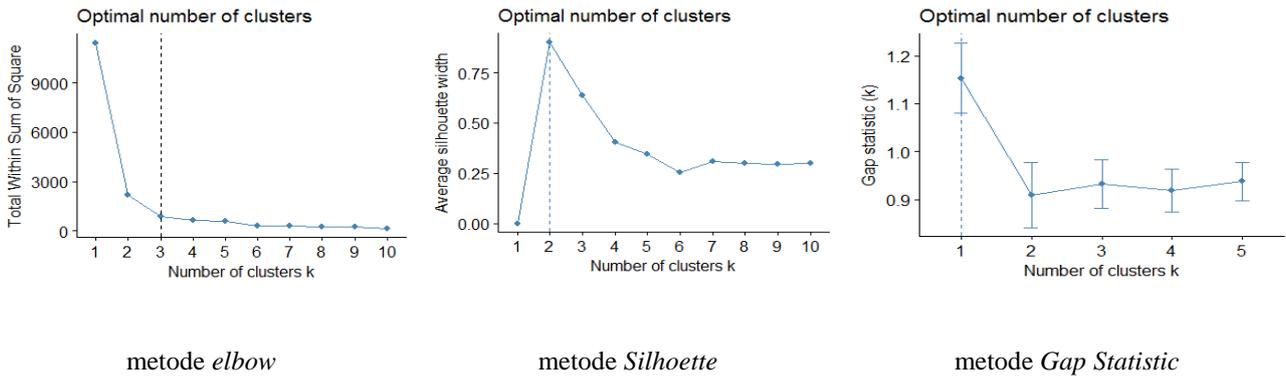
Sumber: hasil pengolahan

Setelah penentuan titik optimal dilakukan, proses dilanjutkan penerapan *clustering* dengan algoritme K-Means. Hasil *clustering* K-Means ditunjukkan pada Tabel 4. Akses internet dengan jumlah *hit* tinggi pada hampir seluruh kategori berada pada *cluster* 3 dengan anggota Fungsional Umum.

Produktivitas dan kinerja pegawai Pusbindiklat Peneliti LIPI berdasarkan perilaku dan pola penggunaan layanan internet

Merujuk penelitian (Fathonah & Hartijasti, 2014), pengukuran produktivitas kerja pegawai diukur melalui level penggunaan akses internet terhadap situs yang tidak mendukung pekerjaan. (Coker, 2011) menemukan bahwa seorang karyawan yang melakukan akses internet terhadap situs yang tidak mendukung pekerjaan yang berlebihan (lebih dari 12% dari total jam kerja) akan mengalami penurunan produktivitas kerja. Berdasarkan rujukan tersebut, pengamatan awal untuk menilai

produktivitas dilakukan dengan melihat frekuensi akses pegawai selama periode bulan Agustus sampai Desember 2016. Profil frekuensi akses dapat dikelompokkan dalam beberapa *cluster*. Dengan kondisi data yang memadai, penentuan jumlah optimal *cluster* dapat dilakukan dengan memanfaatkan 3 metode. Metode pertama menggunakan nilai total wss (*whitin sum square*) atau dikenal dengan istilah metode *elbow*, metode kedua sebagai pembandingan menggunakan metode *silhouette*, untuk menduga kualitas dari *cluster* yang terbentuk, semakin tinggi nilai rata-ratanya maka akan semakin baik, metode ketiga adalah *gap* statistik. Hasil penentuan nilai k optimal dengan menggunakan pendekatan ketiga metode tersebut dapat dilihat pada Gambar 4, nilai k=3 merupakan angka optimal untuk membentuk *cluster*.



Gambar 4. Nilai k optimal dengan 3 metode

Tabel 5. Frekuensi akses internet pegawai Pusbindiklat

Cluster	Level	Jumlah	Persentase
1	Rendah (< 10 hari dalam 5 bulan)	4	8%
2	Sedang (10-66 hari dalam 5 bulan)	17	35%
3	Tinggi (> 66 hari dalam 5 bulan)	28	57%

Sumber: hasil pengolahan

Tabel 6. Waktu akses situs yang tidak mendukung pekerjaan selama jam kerja

Cluster	Kategori	Jumlah	Persentase
1	Rendah, akses situs yang tidak mendukung di tempat kerja (0-4 jam perminggu)	41	84%
2	Sedang, akses situs yang tidak mendukung di tempat kerja (>4-10 jam perminggu)	7	14%
3	Tinggi, akses situs yang tidak mendukung di tempat kerja (>10 jam perminggu)	1	2%

(Sumber level kategori akses: Fathonah dan Hartijasti, 2014)

Hasil *clustering* K-Means dan agregat frekuensi akses pegawai ditunjukkan pada Tabel 5. Presentase frekuensi akses internet dalam 5 bulan terakhir sebagian besar berada pada level tinggi (*cluster* 3) dengan rata-rata frekuensi akses 79 hari dari 105 hari kerja.

Tahap selanjutnya adalah pengamatan terhadap lama waktu akses internet pada kategori situs yang tidak mendukung pekerjaan selama jam kerja. Klasifikasi kategori situs yang dianggap kurang mendukung pekerjaan merujuk pada penelitian (Kim & Byrne, 2011). Hasil *clustering* menggunakan K-Means dengan nilai optimal k=3 membentuk 3 *cluster*. Profil dan karakteristik *cluster* pegawai yang memanfaatkan layanan internet untuk mengakses situs yang tidak

mendukung pekerjaan selama jam kerja tergambar pada Tabel 6.

Tabel 6 menunjukkan produktivitas pegawai diukur berdasarkan lama waktu akses penggunaan internet masih dalam ranah produktif, karena 84% pegawai berada pada *cluster* 1 yang menandakan bahwa umumnya pegawai mengakses situs yang tidak mendukung pekerjaan rata-rata di bawah 4 jam/minggu. Tetapi terdapat satu pengguna yang berada pada *cluster* 3 dengan rata-rata waktu akses mencapai 40 jam/minggu.

Setelah mendapatkan gambaran frekuensi akses dan lama waktu akses internet selama jam kerja, tahap selanjutnya dilakukan identifikasi kategori situs yang paling banyak di akses oleh pegawai. Nilai k=4 adalah nilai optimal untuk membentuk *cluster*

kategori situs yang paling banyak diakses oleh pegawai. Hasil analisis *clustering* dengan K-Means menunjukkan 94% pegawai berada di *cluster 2* seperti ditunjukkan pada Tabel 7.

Tahap akhir dilakukan pemetaan kesesuaian perilaku kerja pegawai dengan pola penggunaan internet yang dilakukannya. Pemetaan dilakukan dengan cara membandingkan jenis kategori situs yang diakses pegawai (Tabel 7) dengan kategori situs yang mendukung pekerjaan berdasarkan tugas dan fungsi jabatannya (Tabel 8).

Hasil pemetaan menunjukkan bahwa pegawai belum sepenuhnya menggunakan internet sesuai dengan tugas dan fungsi jabatannya, seperti terlihat pada Tabel 7 sebanyak 94% pegawai berada pada *cluster 2*, jenis kategori situs pada *cluster* tersebut meliputi *business* (k2), *finance/banking* (k6), *government* (k9), *hacking* (k10), *social network* (k20), *travel* (k21), *uncategory* (k22).

Tabel 7 Cluster kategori situs yang paling banyak diakses oleh pegawai

Cluster	Rendah	Sedang	Tinggi	Sangat tinggi	Jumlah pegawai
1	k1, k2, k3, k4, k6, k8, k10, k11, k16, k18, k21	k9, k15, k17, k20, k22	k5, k12, k14, k19	k13	1 F. Umum
2	k13, k17, k19	k1, k3, k4, k5, k8, k11, k12, k14, k15, k16, k18	k2, k6, k9, k10, k20, k21, k22		1 Peneliti 5 Anpeg 17 F. Umum 5 Perencana 7 Prahum 1 Prakom 9 widyaiswara 1 Honorer
3	k5, k9, k10, k12, k20, k22	k2, k6, k8, k13, k14, k15, k18, k19, k21	k1, k3, k11, k16, k17	k4	1 widyaiswara
4			k4, k10, k17	k1, k2, k3, k5, k6, k8, k9, k11, k12, k13, k14, k15, k16, k18, k19, k20, k21, k22	1 F. Umum

Keterangan kode kategori (k1 sampai k22) dapat dilihat pada Tabel 1, Sumber: hasil pengolahan data

Tabel 8 Kategori situs berdasarkan tugas dan fungsi jabatan

Jabatan	Kategori Situs mendukung pekerjaan	
	Utama	Penunjang
Analisis Kepegawaian	<i>government</i>	<i>education, publication, organization</i>
Perencana	<i>government, education</i>	<i>publication, organization, banking, news</i>
Pranata Humas	<i>government, news</i>	<i>education, publication, organization</i>
Pranata Komputer	<i>ict, education</i>	<i>government, publication, organization, services</i>
Fungsional Umum	<i>government</i>	<i>education, government, organization, publication</i>
Widyaiswara	<i>education, publication</i>	<i>government, organization, news</i>
Peneliti	<i>publication, education</i>	<i>government, organization, news</i>
Analisis Kepegawaian	<i>government</i>	<i>education, publication, organization</i>

Sumber: hasil pengolahan data

Terdapat hasil pemetaan berupa pencilan yang menarik perhatian, salah satunya yaitu satu

pengguna fungsional umum berada pada kategori *cluster 1*. Kondisi tersebut terjadi karena terdapat

akses yang tinggi pada kategori situs *online shopping* (k13). Kondisi lainnya, terdapat pengguna fungsional umum berada pada *cluster* 4. Kondisi tersebut terjadi karena akun salah satu pegawai fungsional umum digunakan oleh beberapa pegawai lainnya. Berikutnya, satu pengguna yang berasal dari jabatan widyaiswara berada pada *cluster* 3. Kondisi tersebut terjadi karena terdapat jumlah *hit* yang tinggi pada kategori *entertainment* (k4).

PENUTUP

Hasil penelitian menunjukkan bahwa proses *web usage mining* dapat diterapkan untuk menganalisis kinerja pegawai di lingkungan Pusbindiklat Peneliti LIPI. Nilai *k* optimal untuk membentuk *cluster* adalah *k*=3 dan *k*=4. Hasil *clustering* dengan *K-Means* menunjukkan aktivitas penggunaan internet dilihat dari frekuensi akses sebagian besar pegawai atau sekitar 57% berada pada level tinggi dengan lama waktu akses internet pada situs yang kurang mendukung pekerjaan umumnya berada pada level rendah atau rata-rata di bawah 4 jam/minggu. Jenis kategori situs yang diakses belum sepenuhnya mendukung pekerjaan. Hasil pemetaan perilaku kerja pegawai dengan pola penggunaan internet masih belum sepenuhnya mendukung tugas dan fungsi jabatannya. Gambaran pola pemanfaatan internet yang dihasilkan dapat dijadikan sebagai salah satu komponen penilaian atau pertimbangan prestasi kerja pegawai oleh pemegang kebijakan. Beberapa saran kebijakan yang dapat diambil antara lain pembatasan akses terhadap kategori situs seperti *social network*, *entertainment*, dan *file sharing*, selanjutnya menghimbau pegawai untuk tidak mengakses situs yang tidak mendukung pekerjaan selama jam kerja. Penataan penggunaan

akun perlu diterapkan, agar pola pemanfaatan internet sebagai salah satu tolak ukur penilaian kinerja dapat lebih terukur.

Selanjutnya perlu dilakukan penelitian lebih lanjut untuk meningkatkan akurasi hitungan waktu atau jam pemakaian internet dengan memperhitungkan waktu penggunaan internet oleh pegawai yang memanfaatkan paket data atau akses internet di luar *proxy*. Selanjutnya untuk menanggulangi tahapan praproses yang terlalu lama perlu dikombinasikan dengan tokenisasi supaya praproses dapat dilakukan lebih cepat. Data penelitian disarankan menggunakan *dataset* terbaru sehingga *monitoring* produktivitas kerja pegawai dapat dilakukan secara berkesinambungan. Kategorisasi situs disarankan dikombinasi dengan tokenisasi supaya praproses dapat dilakukan lebih cepat.

UCAPAN TERIMA KASIH

Ucapan terima kasih disampaikan kepada Lembaga Ilmu Pengetahuan Indonesia dan Pusbindiklat Peneliti LIPI yang telah menyediakan data serta pihak-pihak yang membantu untuk menyelesaikan penelitian ini.

DAFTAR PUSTAKA

- Cadez, I., Heckerman, D., Meek, C., Smyth, P., & White, S. (2003). Model-Based Clustering and Visualization of Navigation Patterns on a Web Site. *Data Mining and Knowledge Discovery*, 7, 399–424.
- Chitraa, V., & Davamani, A. S. (2010). A Survei on Preprocessing Methods for Web Usage Data. *International Journal of Computer Science and Information Security*, 7(3), 78–83. <https://doi.org/10.2200/S00191ED1V01Y2009>

04ICR006

- Chitraa, V., & Thanamani, A. S. (2012). An Enhanced Clustering Technique for Web Usage Mining. *International Journal of Engineering Research & Technology (IJERT)*, 1(4), 1–5.
- Coker, B. L. S. (2011). Freedom to surf: The positive effects of workplace Internet leisure browsing. *New Technology, Work and Employment*, 26(3), 238–247. <https://doi.org/10.1111/j.1468-005X.2011.00272.x>
- Dong, D. (2009). Exploration on Web Usage Mining and Its Application. *Analysis*, 1–4. <https://doi.org/10.1109/IWISA.2009.5072860>
- Fathonah, N., & Hartijasti, Y. (2014). the Influence of Perceived Organizational Injustice Towards Workplace Personal Web Usage and Work Productivity in Indonesia. *South East Asian Journal of Management*, 8(2), 151–166.
- Kerkhofs, J., Vanhoof, K., & Pannemans, D. (2001). Web usage mining on proxy servers: a case study. *Proceedings of Data Mining for Marketing Applications Workshop at ECML/PKDD 2001, September 3-7 2001, Freiburg (Germany)*.
- Kim, S. J., & Byrne, S. (2011). Conceptualizing personal web usage in work contexts: A preliminary framework. *Computers in Human Behavior*, 27(6), 2271–2283. <https://doi.org/10.1016/j.chb.2011.07.006>
- Lüderitz, S. (2006). *Pre-processing of webserver logs for data mining*. Berlin. Diakses dari https://people.cs.kuleuven.be/~bettina.berendt/teaching/2007w/adb/Lecture/OtherSlides/luederitz-presentation1-slides_2006_07_10.pdf
- tanggal 15 September 2016
- Nithya, P., & Sumathi, P. (2012). Novel Pre-Processing Technique for Web Log Mining by Removing Global Noise , Cookies and Web Robots. *International Journal of Computer Applications*, 53(17), 1–6.
- Pamutha, T., Chimphee, S., Kimpan, C., & Sanguansat, P. (2012). Data Preprocessing on Web Server Log Files for Mining Users Access Patterns. *International Journal of Research and Reviews in Wireless Communications (IJRRWC)*, 2(2), 92–98.
- Roiha, N. U. (2017). *Segmentasi Pengguna Web Menggunakan Metode Genetic K-Means Algorithm*. Tesis. Institut Teknologi Sepuluh November.
- Weinreich, H., Obendorf, H., & Herder, E. (2006). Data cleaning methods for client and proxy logs. *Workshop on Logging Traces of Web Activity: The Mechanics of Data Collection; 2006 Mei 23; Edinburgh (GB): Dalhousie University*.
- Xu, J., & Liu, H. (2010). Web User Clustering Analysis Based on K-means Algorithm. *Proceedings of the International Conference on Information Networking and Automation (ICINA)*, 2, V2-6-V2-9. <https://doi.org/10.1109/ICINA.2010.5636772>
- Yusriani, E., & K. Suprpto, Y. (2016). Pemodelan Prediksi Pola Akses Website Pemerintah menggunakan Classification via Regression. *Jurnal Masyarakat Telematika Dan Informasi*, 7(1), 1–12.
- Zhang, Y., Dai, L., & Zhou, Z.-J. (2010). A New Perspective of Web Usage Mining: Using Enterprise Proxy Log. *2010 International*

Conference on Web Information Systems and Mining, 38–42.
<https://doi.org/10.1109/WISM.2010.20>

Zubi, Z. S., Saleh, M., & Raiani, E. (2014). Using Web Logs Dataset via Web Mining for User

Behavior Understanding. *International Journal of Computers and Communications*, 8, 103–111.