

TREN MASA DEPAN PENGGUNAAN APLIKASI WEB USAGE MININGZen Munawar, S.T.,M.Kom¹ ,Novianti Indah Putri,S.T²

1. Manajemen Informatika, Politeknik LP3I Bandung
2. Teknik Informatika, Universitas Bale Bandung

ABSTRACT

Web usage mining is the application of data mining techniques to find usage patterns from web data and to better understand and serve the needs of web-based applications. The web usage mining method uses some background knowledge such as web content, website topology, hierarchy, user navigation and restrictions. In this study the authors propose a new collaborative filtering-based methodology recommendation that deals with excess product problems on large e-commerce sites. The purpose of extracting data is to discover user behavior when accessing the web. This research provides an area of data mining for web usage, theories and future research trends. Web usage mining relates to the technique of extracting log data for user performance used in web-based applications and supports design, e-commerce, services. Web usage mining has three phases, namely preprocessing, pattern detection and pattern learning. The importance of the method of analyzing web data log patterns that have noise and problems when preprocessing. To learn data collection patterns must be professional. Also discussed are various applications for extracting web usage data and existing problems, as well as work related to web usage mining. In this research, we conclude several problems with web usage mining such as finding the desired information, finding related information, learning valuable knowledge, recommendations or personalizing data.

Key Word: Web, web usage mining, web log.

ABSTRAK

Web usage mining adalah penerapan teknik data mining untuk menemukan pola penggunaan dari data web dan untuk memahami dan melayani dengan lebih baik kebutuhan aplikasi berbasis web. Metode web usage mining menggunakan beberapa latar belakang pengetahuan seperti konten web, topologi situs web, Hirarki, navigasi pengguna dan batasan. Dalam penelitian ini penulis mengusulkan rekomendasi berbasis collaborative filtering baru metodologi yang menangani kelebihan produk masalah di situs-situs e-commerce besar. Tujuan dari penggalian data untuk menemukan perilaku pengguna saat mengakses web. Penelitian ini memberikan area penggalian data penggunaan web, teori dan tren penelitian di masa depan. Web usage mining berhubungan dengan teknik penggalian data log untuk kinerja pengguna yang digunakan dalam aplikasi berbasis web serta mendukung perancangan, e-commerce, layanan yang. Web usage mining memiliki tiga fase yaitu preprocessing, deteksi pola dan pembelajaran pola. Pentingnya metode menganalisis pola log data web yang terdapat noise dan permasalahan saat preprocessing. Untuk mempelajari pola pengumpulan data harus secara profesional. Dibahas juga berbagai aplikasi penggalian data penggunaan web dan permasalahan yang ada, serta pekerjaan yang berkaitan web usage mining. Dalam penelitian diperoleh kesimpulan beberapa masalah dengan web usage mining seperti menemukan informasi yang diinginkan, menemukan informasi terkait, mempelajari pengetahuan yang berharga, rekomendasi atau personalisasi data.

Kata Kunci: Web, web usage mining, web log

I. PENDAHULUAN

1.1 Latar Belakang

Situs *web* sangat besar dan sulit untuk diperkirakan pertumbuhan data *web* setiap hari, *web* menyediakan berbagai jenis layanan seperti pemerintah, perdagangan elektronik, berita, dll. *Web usage mining* merupakan pola yang menarik dan berpotensi bermanfaat dari informasi implisit dari *world wide web*. Dukungan *web usage mining* untuk pembuatan desain situs *web*; menyediakan personalisasi server *web* dan keputusan bisnis lainnya. *Web usage mining* terdiri dari empat langkah dan terdiri dari tiga tahap, yaitu pra-pemrosesan, penemuan pola, dan analisis pola. Langkah pertama adalah pengumpulan data, langkah kedua adalah preprocessing, langkah ketiga adalah penemuan pola dan langkah terakhir adalah analisis pola. *Web usage mining* adalah aplikasi *data mining* : Informasi struktur hyper-link dari suatu situs *web* yaitu, berbagai tag HTML yang digunakan untuk menautkan satu halaman ke halaman lain dan satu situs *web* ke yang lain. Penggunaan *web*: Informasi yang mencerminkan penggunaan sumber daya *Web*, misalnya., entri dalam riwayat browser *web* dan file sementara Internet, server proxy dan log server *web*.

Kemudahan dan kecepatan transaksi bisnis dapat dilakukan dilakukan melalui *web* telah menjadi kekuatan pendorong utama dalam pertumbuhan cepat perdagangan elektronik. khususnya, e-aktivitas perdagangan yang melibatkan pengguna akhir sedang berlangsung sebuah revolusi yang signifikan. Kemampuan untuk melacak penelusuran pengguna perilaku ke masing-masing klik *mouse* telah membawa vendor dan konsumen akhir lebih dekat dari sebelumnya. Sekarang mungkin bagi vendor untuk mempersonalisasikan pesan produknya pelanggan individu dalam skala besar, sebuah fenomena yang sedang disebut sebagai kustomisasi massal. Skenario yang dijelaskan di atas adalah salah satu dari banyak aplikasi yang mungkin *web usage mining*, yang merupakan proses penerapan teknik *data mining* untuk penemuan pola penggunaan dari data *web*, ditargetkan ke berbagai aplikasi. Upaya *data mining* yang terkait dengan *web*, yang disebut *web mining*, dapat secara luas dibagi menjadi tiga kelas, misalnya. konten *mining*, *web usage*, dan struktur *mining*. Struktur *web* eb proyek *mining* seperti [1][2] dan konten *web mining* proyek-proyek seperti [3][4] berada di luar cakupan survei ini.

Bulent Ozel et al., [5] mengusulkan tautan *web hybrid* ramalan. Tumbuhnya *web* meningkatkan navigasi masalah. Menggunakan penggunaan *web* dapat memprediksi perilaku pengguna dan membantu perancang untuk meningkatkan desain agar menarik penggunaan

pengguna. Dalam tulisan ini penulis menggunakan keduanya teknik *data mining* aturan asosiasi dan rantai Markov model. Proses cluster tautan *hybrid* serupa halaman untuk meningkatkan efisiensi model yang diusulkan.

Yoon Ho Cho et al., [6] menyajikan pertumbuhan e-commerce yang cepat menunjukkan permintaan pelanggan di *web*. Di dalam kertas e-commerce menyebabkan kelebihan pelanggan *web* dalam waktu yang lama.

Untuk mengatasi kelebihan ini, beragam metode telah dilakukan bekas. Pemfilteran Kolaboratif adalah rekomendasi metode telah mengatasi batasan pada metode yang ada. Rana [7] memfokuskan pada teknik yang dapat memprediksi perilaku pengguna saat berinteraksi dengan situs *web*. Pendekatannya multidisiplin itu *web* terus berkembang. Sehingga bisa diprediksi pengguna perlu membuatnya kuat, terukur, dan efisien rancangan. Data yang dihasilkan oleh sesi atau pengguna surfer perilaku. Mereka membahas alat yang tersedia di Internet aplikasi penggunaan *web mining* dan diakhiri dengan tantangan dan tren masa depan dalam penelitian. Zaiane [8] mengembangkan penemuan pengetahuan *WeblogMiner* alat dari *file log* server. Dengan demikian dapat meningkatkan kinerja sistem, dan meningkatkan kualitas situs *web* dan mengirimkan data yang bermanfaat kepada pengguna akhir.

1.2 Kajian Teoritis dan Ilmiah

Sebuah arsitektur pandangan holistik, ditambah dengan perspektif konstruktivis yang kuat (Thórisson, 2012).

1.3 Pemrosesan Real Time

Meskipun jelas, penting untuk diingat bahwa manusia adalah informasi sistem pengolahan waktu-nyata. Tidak ada pilihan bagi kami untuk menunda waktu dan keluar jalur untuk berunding tentang dunia nyata yang bergerak, apakah kita suka atau tidak.

II. KAJIAN TEORITIS

2.1. Survei

Sebuah Studi - Alat Penelitian *web usage mining* diberikan oleh Chhavi Rana [9] mempresentasikan persyaratan dunia saat ini dan mendaftar beberapa tantangan dan masalah tren masa depan dari pengguna yang mendapat perhatian situs *web* berkualitas. Abraham [10] telah menggabungkan *data mining* dan World Wide Web untuk penelitian. Itu penemuan pengetahuan bisa berusaha untuk mendapatkan dari internet data sekunder. Penggunaan alat bantu *i-miner* untuk mengoptimalkan arsitektur bersamaan dari algoritma *fuzzy clustering* untuk menemukan

kluster informasi untuk dianalisis tren inferensi fuzzy digunakan. Cooley [11] menggambarkan penggunaan pola *web usage mining* bisa ditemukan dari urutan yang tidak menarik. Beberapa upaya penelitian mengandalkan aturan yang tidak menarik. Itu Penggunaan sistem informasi situs *web filter* konten *web* dan informasi struktur web dari situs web untuk mengidentifikasi pola *mining* yang menarik set item yang sering dari dunia nyata. *Web usage mining* adalah aplikasi teknik *data mining* untuk ditemukan pola-pola dari *web*, untuk memahami berbasis aplikasi *web*. Data di *web* dapat digali dari data sekunder di *web*. Kita bisa mengklasifikasikan data yang berada di *web* [12]. Shahabi [13] Melacak interaksi pengguna dengan *web* dan menghasilkan pengguna profil secara *real time* dengan menggunakan Konektivitas Matriks Model (CM-Model), yang menunjukkan efisiensi dan skalabilitas atribut partisipatif pengguna di *web* untuk memvisualisasikan jalur navigasi pengguna secara *real time*. Dengan menggunakan grafik data penggunaan *web* bisa mewakili [14].

Metode *web usage mining* menggunakan beberapa latar belakang pengetahuan seperti konten web, topologi situs *web*, Hirarki, navigasi pengguna dan batasan. dalam makalah ini penulis mengusulkan rekomendasi berbasis *collaborative filtering* baru metodologi yang menangani kelebihan produk masalah di situs-situs e-commerce besar [15]. Penggunaan Kerangka Kerja *web mining* untuk Profil Pengguna yang Sedang *Evolving mining* di Situs *web* Dinamis oleh Olfa Nasraoui [16] mempresentasikan kerangka kerja lengkap untuk menemukan pengguna profil untuk menemukan pola dari file log situs *web* asli. Minat berbasis pengguna dianalisis pada situs *web* dengan kueri penelusuran untuk mengekstrak data log *web*. L.Chen [17] mengembangkan *webmate*; sebagai agen proxy membantu pengguna sebagai penjelajahan dan penelusuran *web* yang efektif. Eksperimen menggambarkan yang sistem yang ada yang diusulkan dari profil pengguna yang digunakan dalam berbagai sudut pandang bisnis dan aplikasi. Survei di *web usage mining* telah dilakukan oleh Koutri, Avouris, dan Daskalaki [18]. Membangun model pengguna dengan menggunakan Teknik *WebLogMiner* digunakan untuk mengungkap yang tersembunyi pola dalam web. Informasi akses web disimpan dalam sumber data. Penggunaan teknik untuk membangun situs *web* adaptif. Eirinaki M. [19] membahas kebutuhan pengguna dengan analisis perilaku navigasi di *web*. Untuk personalisasi

yang dikembangkan penulis sebagai sistem itu membuat baik log penggunaan *web* dan konten situs web itu merangkum penemuan pengetahuan dari tautan semantik. Buchner [20] memperkenalkan algoritma baru di berbagai *web* dapat menemukan penemuan berurutan tradisional. Dengan demikian menunjukkan fungsi serta skalabilitas. Ling [21] pemasaran langsung adalah proses mengidentifikasi produk dibeli siapa dan yang dijual. Alat *data mining* digunakan untuk pemasaran langsung. Informasi terkait web adalah target yang tepat dan populer untuk penemuan pengetahuan, penemuan pengetahuan menyangkut proses konten web, struktur web dan web pemakaian. Dari minat pengguna situs web berkualitas dapat mengembangkan dan menemukan aturan dan pola untuk diambil beberapa langkah menarik [22]. Senkul dan Salin [23] menghasilkan teknik *web usage mining* untuk *web* restrukturisasi dan rekomendasi lokasi. Menyelidiki data semantik dari halaman web di Internet pola dihasilkan untuk urutan yang sering. Pola navigasi pengguna yang sering diukur dengan mekanisme yang melibatkan rekomendasi halaman web. SpeedTracer [24] mengembangkan perilaku berselancar pengguna oleh *file log server*. Sesi pengguna oleh merekonstruksi jalur traversal melalui halaman pengarah dan dihasilkan tiga jenis laporan yang disiapkan. Rao, Kumari, dan Raju, [25] mengembangkan aturan asosiasi *mining* dengan metode tambahan. Dengan menggunakan algoritma ini agar sesuai dengan skenario log perubahan dinamis dengan teknik tambahan. Teknik ini lebih efisien untuk menjalankan sejumlah database. Ujwala Patil [26] mengusulkan survei atau prediksi permintaan di masa mendatang ekstraksi dari *file log web*. *Web file log* sangat berguna untuk memprediksi perilaku pengguna dengan berbagai cara. Menyediakan data masa lalu, evaluasi saat ini dan pembaruan *web usage mining*. Ekstraksi Aturan Bisnis dari log Web ke Meningkatkan Penggunaan *Web Mining* diberikan oleh Sawan Bhawaret al., [27]. merepresentasikan data otomatis diekstraksi dengan permintaan, pengorganisasian, dan analisis. Ekstraksi Informasi adalah tugas ekstraksi terstruktur informasi dari yang tidak terstruktur. *Input* adalah *file log* dan mengekstraksi informasi dari sesi yang dilakukan oleh pengguna. Ini dapat meningkatkan penggunaan *web*, mengusulkan situs web dengan halaman *web* terbatas bisa mudah memprediksi permintaan pengguna dari *web* tertentu situs.

2.2. Web data

Salah satu langkah kunci dalam *Knowledge Discovery in Databases* [28] adalah membuat target data yang sesuai untuk tugas-tugas *data mining*.

Pada *web mining*, data dapat dikumpulkan di sisi server, sisi klien, server *proxy*, atau diperoleh dari database organisasi (yang berisi data bisnis atau data *web* terkonsolidasi). Setiap jenis pengumpulan data berbeda tidak hanya dalam hal lokasi sumber data, tetapi juga jenis data yang tersedia, segmen populasi dari mana data dikumpulkan, dan metode implementasinya. Ada banyak jenis data yang dapat digunakan di *web mining*. Penelitian ini mengklasifikasikan data tersebut ke dalam jenis-jenis berikut: Konten: Data nyata di halaman *web*, yaitu data halaman *web* yang dirancang untuk disampaikan kepada pengguna. Ini biasanya terdiri dari, tetapi tidak terbatas pada, teks dan grafik. Struktur: Data yang menggambarkan pengorganisasian konten. Informasi struktur intra-halaman mencakup pengaturan berbagai tag HTML atau XML dalam halaman tertentu. Hal ini dapat direpresentasikan sebagai struktur pohon, di mana tag *html* menjadi root dari pohon. Jenis utama dari informasi struktur antar halaman adalah hyper-link yang menghubungkan satu halaman ke halaman lainnya. Penggunaan: Data yang menjelaskan pola penggunaan halaman *web*, seperti alamat IP, referensi halaman, dan tanggal serta waktu akses. Profil Pengguna: Data yang menyediakan informasi demografis tentang pengguna situs Web. Ini termasuk data pendaftaran dan informasi profil pelanggan.

Data penggunaan yang dikumpulkan di berbagai sumber akan mewakili pola navigasi segmen berbeda dari keseluruhan jejak *web*, mulai dari perilaku penelusuran satu-pengguna, satu-situs hingga pola akses multi-pengguna, multi-situs.

Log server web adalah sumber penting untuk melakukan *web usage mining* karena secara eksplisit mencatat perilaku penelusuran pengunjung situs. Data yang dicatat dalam log server akan memperbaiki akses (mungkin bersamaan) dari situs *web* oleh banyak pengguna. *File log* ini dapat disimpan dalam berbagai format seperti format Common log atau Extended log. Namun, data penggunaan situs yang direkam oleh *log server* mungkin tidak sepenuhnya dapat diandalkan karena adanya berbagai tingkat *caching* di lingkungan Web. Tampilan halaman dalam *cache* tidak direkam dalam *log server*. Selain itu, informasi penting apa pun yang dilewatkan melalui metode POST tidak

akan tersedia dalam *log server*. Teknologi *packet sniffing* adalah metode alternatif untuk mengumpulkan data penggunaan melalui log server. Paket *sniffer* memantau lalu lintas jaringan yang datang ke server Web dan mengekstrak data penggunaan langsung dari paket TCP / IP. *Server web* juga dapat menyimpan jenis penggunaan informasi lainnya informasi seperti cookie dan data permintaan dalam log terpisah. *Cookie* adalah token yang dihasilkan oleh *server web* untuk masing-masing browser klien untuk secara otomatis melacak pengunjung situs. Pelacakan pengguna individu bukanlah tugas yang mudah karena model koneksi tanpa kewarganegaraan dari protokol HTTP. *Cookie* bergantung pada kerja sama pengguna implisit dan karenanya telah meningkatkan kekhawatiran tentang privasi pengguna. Data kueri juga biasanya dihasilkan oleh pengunjung online sambil mencari halaman yang relevan dengan kebutuhan informasinya. Selain data penggunaan, sisi server juga menyediakan data konten, informasi struktur, dan meta-informasi halaman *web* (seperti ukuran file dan waktu modifikasi terakhirnya). *Server web* juga mengandalkan utilitas lain seperti skrip CGI untuk menangani data yang dikirim kembali dari browser klien. *Server web* yang menerapkan standar CGI mengurai URI 1 dari *file* yang diminta untuk menentukan apakah itu adalah program aplikasi. URI untuk program CGI dapat berisi nilai parameter tambahan untuk diteruskan ke aplikasi CGI. Setelah program CGI menyelesaikan eksekusi, *server web* mengirim output aplikasi CGI kembali ke browser.

Uniform Resource Identifier (URI) adalah definisi yang lebih umum yang mencakup *Uniform Resource Locator* (URL). Pengumpulan data sisi klien dapat diimplementasikan dengan menggunakan agen jarak jauh (seperti *Javascripts* atau *applet Java*) atau dengan memodifikasi kode sumber *browser* yang ada (seperti *Chrome* atau *Mozilla*) untuk meningkatkan kemampuan pengumpulan datanya. Implementasi metode pengumpulan data sisi klien membutuhkan kerja sama pengguna, baik dalam mengaktifkan fungsionalitas *Javascripts* dan *applet Java*, atau untuk secara sukarela menggunakan browser yang dimodifikasi. Pengumpulan sisi klien memiliki keuntungan lebih dari pengumpulan sisi server karena hal itu memperbaiki masalah keduanya *caching* dan sesi identifikasi. Namun, *applet Java* berkinerja tidak lebih baik dari server log dalam hal menentukan waktu tampilan halaman yang sebenarnya. Bahkan, hal itu mungkin menimbulkan beberapa *overhead* tambahan terutama ketika *applet Java* dimuat untuk pertama kalinya.

Javascrpts, di sisi lain, mengkonsumsi sedikit waktu interpretasi tetapi tidak dapat menangkap semua klik pengguna (seperti memuat ulang atau tombol kembali). Metode ini hanya akan mengumpulkan perilaku menjelajah satu pengguna, satu situs. Peramban yang dimodifikasi jauh lebih fleksibel dan akan memungkinkan pengumpulan data tentang satu pengguna melalui beberapa situs web. Bagian tersulit dalam menggunakan metode ini adalah meyakinkan pengguna untuk menggunakan browser untuk aktivitas *browsing* harian. Hal ini dapat dilakukan dengan menawarkan insentif kepada pengguna yang bersedia menggunakan browser, mirip dengan program insentif yang ditawarkan oleh perusahaan seperti NetZero [29] dan All Advantage [30] yang memberikan penghargaan kepada pengguna karena mengklik banner.

Proxy web bertindak sebagai cache tingkat menengah antara browser klien dan *server web*. *Caching proxy* dapat digunakan untuk mengurangi waktu pemuatan halaman *web* yang dialami oleh pengguna serta beban lalu lintas jaringan di sisi server dan klien [31]. Kinerja *cache proxy* tergantung pada kemampuan mereka untuk memprediksi permintaan halaman di masa depan dengan benar. Jejak *proxy* dapat mengungkapkan permintaan HTTP aktual dari banyak klien ke beberapa *server web*. Berfungsi sebagai sumber data untuk mengkarakterisasi perilaku penelusuran sekelompok pengguna anonim yang berbagi *server proxy* umum.

III. WEB USAGE MINING

Web usage mining adalah cabang *web mining*, yang merupakan divisi dari Data Mining. Metode *data mining* yang berharga dan informasi yang berharga dari basis data yang sangat besar bernama *data mining*. Ekstraksi atribut *web usage mining*, penggunaan layanan *web* klien. Catatan yang dicapai kemudian dapat berguna dalam berbagai hal pendekatan misalnya, pemeriksaan esensi [32]. *Web usage mining* yang baik sebagai konstituen intelijen bisnis besar dalam bisnis. Itu berlaku untuk memilih kemajuan bisnis melalui penggunaan layanan *web* yang mahir. Hal ini sangat penting untuk hubungan manajemen pelanggan karena menjamin kinerja klien hingga antarmuka antar organisasi dan klien [33].

Nama lain dari *web usage mining* disebut juga *web log mining* yang mempunyai arti aplikasi metode data mining pada area penyimpanan *log web* besar untuk menentukan nilai pengetahuan tentang pola perilaku pelanggan dan informasi penggunaan situs web untuk memfasilitasi

berbagai bidang desain situs *web*. Sumber daya utama statistik untuk *web usage mining* terdiri dari konten *log* bersama dengan beberapa server web yang tak tertandingi di dunia. Ada tiga tahap dalam *web usage mining*. Data *log* pengguna secara kolektif mulai berbeda, sumber daya sebagai server alternatif, sisi klien dan server sisi sebagainya. Pengolah Data: membuat suksepsi pemrosesan *file log web* yang mencakup catatan pembersihan, pengenalan sesi, pengakuan pelanggan, jalur posisi akhir dan deteksi transaksi.

Deteksi pola: prinsip data yang berbeda metode *mining* untuk memproses data yang terkait dengan statistik. studi, asosiasi, pengelompokan, dan pola selanjutnya dan seterusnya. Pembelajaran pola: pola sebelumnya ditemukan pada *log web*, konvensi yang membosankan disaring. Pembelajaran dilengkapi dengan sistem penyelidikan informasi untuk kubus data kejadian atau SQL untuk menyajikan tindakan OLAP (*On-Line Analytical Processing*). Seluruh tiga tahap mewakili melalui diagram berikut. Tujuan dari penelitian ini adalah untuk menyajikan studi *web usage mining* bersama dengan analisis tiga tahap. Sektor pengumpulan data mencatat sumber daya data yang berbeda dan berbagai bidang fungsi atau aplikasi di mana *web usage mining* sesuai [34].

3.1 Preprocessing

Preprocessing terdiri dari mengonversi penggunaan, konten, dan struktur informasi yang terkandung dalam berbagai data yang tersedia sumber ke dalam abstraksi data yang diperlukan untuk penemuan pola. Penggunaan *preprocessing* bisa dibilang tugas yang paling sulit di proses *web usage mining* karena ketidaklengkapan data yang tersedia. Kecuali mekanisme pelacakan sisi klien digunakan, hanya alamat IP, agen, dan klik sisi server streaming tersedia untuk mengidentifikasi sesi pengguna dan server.

Beberapa masalah yang biasanya ditemui adalah: Alamat IP tunggal / Beberapa Sesi Server – Internet penyedia layanan (ISP) biasanya memiliki kumpulan *proxy server* yang diakses pengguna melalui *web*. *Server proxy* tunggal mungkin memiliki beberapa pengguna yang mengakses situs *web*, berpotensi selama periode waktu yang sama. Beberapa alamat IP / Sesi Server Tunggal - Beberapa ISP atau alat privasi menetapkan secara acak setiap permintaan dari pengguna ke salah satu dari beberapa alamat IP. Dalam hal ini, sesi server tunggal dapat memiliki beberapa alamat IP.

Multiple IP address / Single User - Seorang pengguna yang mengakses Web dari mesin yang berbeda akan memiliki perbedaan Alamat IP dari sesi ke sesi. Ini membuat *track* dengan melakukan kunjungan berulang dari kesulitan pengguna yang sama. *Multiple Agent / Single User* - Sekali lagi, pengguna yang menggunakan lebih dari satu *browser*, bahkan pada mesin yang sama, akan muncul sebagai banyak pengguna. Dengan asumsi setiap pengguna sekarang telah diidentifikasi

(melalui *cookies*, *login*, atau *IP / agent / path analysis*), aliran klik untuk setiap pengguna harus dibagi menjadi beberapa sesi. Karena permintaan halaman dari *server* lain biasanya tidak tersedia, ini menjadi sulit untuk mengetahui kapan pengguna telah meninggalkan situs *web*. Tiga puluh menit *timeout* sering digunakan sebagai metode standar untuk memecahkan klik-*stream* pengguna ke dalam sesi. Batas waktu tiga puluh menit didasarkan pada hasil [35]. Ketika ID sesi dikeluarkan tidur di setiap URI, definisi sesi diatur oleh server konten.

Sementara konten yang tepat disajikan sebagai hasil dari setiap pengguna sering tersedia dari permintaan di *server log*, terkadang perlu memiliki akses ke konten informasi *server* juga. Karena konten *server* dapat berupa variabel keadaan untuk setiap sesi aktif, informasi diperlukan untuk menentukan dengan tepat konten apa yang dilayani oleh permintaan pengguna tidak selalu tersedia di URI. Masalah yang dihadapi saat praproses data penggunaan adalah itu menyimpulkan referensi halaman yang di-*cache*. Satu-satunya metode yang dapat diverifikasi untuk melacak tampilan halaman dalam *cache* adalah untuk memantau penggunaan dari sisi klien. Bidang pengarah untuk setiap permintaan dapat digunakan untuk mendeteksi beberapa contoh ketika halaman dalam *cache* telah dilihat.

3.2 Pemrosesan Ulang Konten

Preprocessing konten terdiri dari konversi teks, perbaikan umur, skrip, dan *file* lain seperti multimedia ke dalam bentuk yang berguna untuk proses *web usage mining*. Sering, hal ini terdiri dari melakukan konten *mining* seperti klasifikasi atau pengelompokan. Saat menerapkan *data mining* ke konten situs *web* adalah bidang penelitian yang menarik di dalamnya, dalam konteks konten *web usage mining* dari suatu situs dapat digunakan untuk menyaring input, atau output dari algoritma penemuan pola. Misalnya, hasil algoritma klasifikasi dapat digunakan untuk membatasi diskusi ered pola untuk orang-orang yang berisi tampilan halaman tentang tertentu subjek atau kelas produk. Selain mengklasifikasikan atau mengelompokkan tampilan halaman berdasarkan topik, tampilan halaman juga bisa diklasifikasi sesuai dengan tujuan penggunaannya [36]. Tampilan halaman dapat dimaksudkan untuk menyampaikan informasi (melalui teks, grafik, atau multimedia lainnya), kumpulkan informasi dari pengguna, izinkan navigasi (melalui daftar tautan *hypertext*), atau beberapa penggunaan kombinasi. Tujuan penggunaan tampilan halaman dapat juga menyaring sesi sebelum atau setelah penemuan pola.

Untuk menjalankan algoritma *mining* konten pada tampilan halaman, informasi pertama-tama harus diubah menjadi kemampuan

kuantitatif format. Beberapa versi model ruang vektor [37] diketikkan digunakan untuk mencapai ini. File teks dapat dipecah menjadi vektor kata. Kata kunci atau deskripsi teks bisa diganti untuk grafis atau multimedia. Isi dari tampilan halaman statis dapat dengan mudah diproses sebelumnya dengan mem-parsing HTML dan format ulang informasi atau jalankan algoritma tambahan seperti yang diinginkan. Tampilan halaman dinamis hadir lebih merupakan tantangan. *Server* konten yang menggunakan personal-teknik dan / atau memanfaatkan basis data untuk dibangun tampilan halaman mungkin dapat membentuk lebih banyak tampilan halaman dari yang bisa diproses secara praktis. Satu set *server* yang diberikan sesi hanya dapat mengakses sebagian dari tampilan halaman yang mungkin untuk situs dinamis besar. Juga konten dapat direvisi secara teratur. Konten dari setiap tampilan halaman menjadi diproses harus \ dirakit, baik dengan permintaan HTTP dari *browser*, atau kombinasi templat, skrip, dan akses basis data. Jika hanya sebagian halaman yang melihatnya diakses sudah diproses sebelumnya, *output* dari setiap klasifikasi algoritma atau *clustering* mungkin miring. *Preprocessing* struktur, struktur situs dibuat oleh tautan *hypertext* menjadi *tween page views*. Struktur dapat diperoleh dan diproses dengan cara yang sama seperti konten situs. Konten dinamis menimbulkan lebih banyak masalah dari tampilan halaman statis. Struktur situs yang berbeda mungkin memiliki dan akan dibangun untuk setiap *server session*.

3.3 Penemuan Pola

Penemuan pola mengacu pada metode dan algoritma dikembangkan dari beberapa bidang seperti statistik, *data mining*, *machine learning* dan pengenalan pola. Namun demikian bukan maksud dari makalah ini untuk menggambarkan semua algoritma dan teknik yang berasal dari bidang ini. [38]. Menggambarkan jenis kegiatan *mining* yang telah dilakukan diterapkan ke *web domain*. Metode dikembangkan dari bidang yang lain harus mempertimbangkan jenis data yang berbeda abstraksi dan pengetahuan sebelumnya yang tersedia untuk *web mining*.

Sebagai contoh, dalam penemuan aturan asosiasi, gagasan tentang transaksi untuk analisis keranjang pasar tidak memperhitungkan mempertimbangkan urutan pemilihan item. Bagaimanapun pernah, di *web usage mining*, sesi *server* adalah perintah urutan halaman yang diminta oleh pengguna. Selanjutnya, karena kesulitan dalam mengidentifikasi sesi unik, tambahan sebelumnya diperlukan pengetahuan.

Analisis statistik, teknik statistik adalah metode yang paling umum untuk menyalurkan pengetahuan tentang pengunjung ke situs *web*. dengan menganalisis pada sesi sesi, seseorang dapat melakukan berbagai macam deskripsi analisis statistik (frekuensi, rata-rata, median, dan lain-lain) pada variabel seperti tampilan halaman,

waktu menonton dan panjang jalur navigasi. Banyak alat analisis lalu lintas *web* menghasilkan laporan berkala yang berisi informasi statistik seperti halaman yang paling sering diakses, waktu tampilan rata-rata halaman atau panjang rata-rata jalur melalui situs. Laporan dapat mencakup analisis kesalahan tingkat rendah terbatas seperti titik masuk tidak sah atau menemukan yang paling umum URI tidak valid. Meski kurang dalam analisisnya, tipe pengetahuan ini dapat berguna untuk meningkatkan kinerja sistem, meningkatkan keamanan sistem, memfasilitasi tugas modifikasi situs, dan menyediakan dukungan untuk keputusan pemasaran. Aturan asosiasi, pembuatan aturan asosiasi dapat digunakan untuk menghubungkan halaman itu paling sering dirujuk bersama dalam satu sesi server tunggal.

Dalam konteks *web usage mining*, aturan asosiasi merujuk ke set halaman yang diakses bersama dengan dukungan nilai melebihi beberapa ambang batas yang ditentukan. Halaman-halaman ini mungkin tidak terhubung langsung satu sama lain melalui hyperlink. Misalnya, untuk penemuan aturan asosiasi menggunakan algoritma Apriori [39] atau salah satu variannya dapat mengungkapkan korelasi antara pengguna yang mengunjungi halaman yang berisi produk elektronik bagi mereka yang mengakses halaman tentang peralatan olahraga.

Selain berlaku untuk aplikasi bisnis dan pemasaran Jika ada, ada atau tidak adanya aturan semacam itu dapat membantu desainer *web* untuk merestrukturisasi situs *web*. Aturan asosiasi juga dapat berfungsi sebagai heuristik untuk pengambilan dokumen sebelumnya untuk mengurangi latensi yang dirasakan pengguna saat memuat halaman dari situs jarak jauh. *Clustering* adalah teknik untuk mengelompokkan satu set item memiliki karakteristik yang serupa. Di domain penggunaan *web*, ada dua jenis cluster yang menarik untuk ditemukan: kelompok penggunaan dan kelompok laman. Pengelompokan pengguna cenderung untuk membentuk kelompok pengguna yang memperlihatkan paten penelusuran serupa terns. Pengetahuan seperti itu sangat berguna untuk menyimpulkan pengguna demografis untuk melakukan segmentasi pasar.

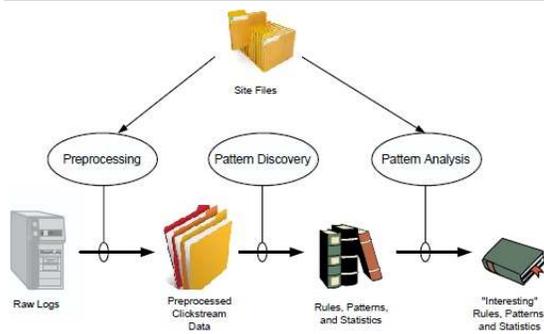
Aplikasi e-commerce menyediakan konten *web* yang dipersonalisasi untuk pengguna. Di sisi lain, pengelompokan halaman akan menemukan grup halaman yang memiliki konten terkait. Ini informasi berguna untuk mesin pencari Internet dan *web* penyedia bantuan. Baik dalam aplikasi, permanen atau halaman HTML dinamis dapat dibuat yang menyarankan terkait hyperlink ke pengguna sesuai dengan kueri atau masa lalu pengguna sejarah kebutuhan informasi.

Klasifikasi adalah tugas memetakan item data menjadi satu dari beberapa kelas yang telah ditentukan [40]. Di domain *web*, satu tertarik untuk mengembangkan profil pengguna milik kelas atau

kategori tertentu. Ini membutuhkan ekstraksi dan pemilihan fitur yang paling menggambarkan properti diberikan kelas atau kategori. Klasifikasi dapat dilakukan dengan menggunakan algoritma pembelajaran induktif yang diawasi seperti keputusan klasifikasi pohon, klasifikasi Bayesian yang naif, k-tetangga terdekat klasifikasi, Mendukung Mesin Vektor dan lain-lain. Misalnya, klasifikasi pada *log server* dapat mengarah pada penemuan informasi aturan seperti: 30% pengguna yang menempatkan *online* urutan / produk / musik berada di kelompok usia 18-2. Teknik penemuan pola sekuensial mencoba dan pola-pola antar-sesi sedemikian rupa sehingga kehadiran satu set item diikuti oleh item lain dalam urutan ses- yang dipesan waktu atau episode. Dengan menggunakan pendekatan ini, pemasar *web* dapat memperkirakan pola kunjungan di masa mendatang yang akan membantu di menempatkan iklan yang ditujukan untuk grup pengguna tertentu. Jenis analisis temporal yang dapat dilakukan secara berurutan pola awal termasuk analisis tren, deteksi titik perubahan, atau analisis kesamaan.

Pemodelan ketergantungan adalah penemuan pola lain yang bermanfaat tugas di *web mining*. Tujuannya di sini adalah untuk mengembangkan model mampu mewakili ketergantungan yang signifikan antara berbagai variabel dalam domain *web*. Sebagai satu contoh, mungkin tertarik untuk membangun model yang mewakili perbedaan tahapan yang dialami pengunjung saat berbelanja di toko online berdasarkan tindakan yang dipilih (misalnya, dari pengunjung biasa ke pembeli potensial). Ada beberapa pembelajaran probabilistic Teknik yang dapat digunakan untuk memodelkan penelusuran perilaku pengguna. Teknik-teknik tersebut termasuk Hidden Markov Model dan Bayesian Belief Networks. Pemodelan *web* pola umur tidak hanya menyediakan kerangka teori untuk menganalisis perilaku pengguna tetapi berpotensi bermanfaat untuk memprediksi konsumsi sumber daya *web* masa depan. Informasi seperti dapat membantu mengembangkan strategi untuk meningkatkan penjualan produk-produk yang ditawarkan oleh situs *web* atau meningkatkan navigasi kenyamanan pengguna.

Analisis pola adalah langkah terakhir dalam penggunaan *web* secara keseluruhan proses *mining* seperti yang dijelaskan pada gambar 1. Motivasi di balik analisis pola adalah menyaring aturan yang tidak menarik atau pola dari himpunan ditemukan dalam fase penemuan pola.



Gambar 1. Proses *web usage mining* tingkat tinggi

Metodologi analisis yang tepat biasanya diatur oleh aplikasi yang melakukan *web mining*. Bentuk analisis pola terdiri dari kueri pengetahuan mekanisme seperti SQL. Metode lain adalah memuat penggunaan data ke dalam kubus data untuk melakukan operasi OLAP. Teknik visualisasi, seperti pola grafik atau menandatangani warna dengan nilai yang berbeda, sering dapat menyorot keseluruhan pola atau tren dalam data. Konten dan informasi struktur mation dapat digunakan untuk menyaring pola yang berisi halaman dari jenis penggunaan tertentu, jenis konten, atau halaman yang cocok struktur hyperlink tertentu.

IV. APLIKASI WEB USAGE MINING

Area aplikasi *web usage mining* bisa berupa aplikasi reorganisasi situs, prediksi tautan web, personalisasi dan pra-pengambilan web. Dua area utama adalah topik yang menarik. Dua berikutnya menggunakan area dapat dievaluasi dalam konteks e-commerce.

4.1 Meningkatkan Kinerja Sistem

Kinerja layanan web adalah masalah signifikan untuk kepuasan pengguna. *web usage mining* adalah area penelitian penting untuk menemukan lalu lintas web prosedur, yang dapat digunakan untuk memperkuat yang baru kebijakan untuk meningkatkan konser *server web* [41]. Penyeimbangan beban, transmisi metode caching web atau distribusi akun adalah area aplikasi umum web mining untuk peningkatan kinerja [42]. Mengharapkan permintaan halaman *web* berikutnya untuk pengguna yang menuntut koleksi adalah masalah penelitian penting dalam hal ini lingkungan Hidup. Proses ini sangat berharga terutama di layanan *web* menggunakan statis atau dinamis dengan mengurangi kegunaan *caching web* pada pengguna dan tingkat server. Dengan *proxy* dalam klasifikasi untuk pra-pengambilan halaman juga merupakan metode lain untuk kinerja peningkatan [43][44]. Restrukturisasi situs. asosiasi tautan dan komposisi konten situs web apa pun adalah dua faktor utama situs *web* mana pun. Itu tren saat ini dalam alat *web mining* lebih pendek urutan

navigasi, untuk mendukung motif itu aksesibilitas halaman objektif di beberapa domain *web* perlu ditingkatkan [45]. Situs restrukturisasi topologi dari setiap bidang web dapat mencapai ini. Restrukturisasi dapat dilakukan sehubungan dengan pola yang sering tidak diperhatikan di akhir *web usage mining*. Data penggunaan web memberikan informasi bersama dengan desain situs *web* apa pun yang menarik untuk aktivitas pengguna [46]. Untuk menampilkan momen tinggal di halaman memberi halaman, yang tidak menarik. Vendor dari situs web dapat membangun kembali halaman-halaman ini dan mengamati perilaku pengguna di halaman ini. Dua teknik organisasi konten dan struktur ke situs web adaptif. Model diberikan dalam perubahan organisasi situs *web* sehubungan dengan pola penggunaan terbuka [47] [48].

4.2 Personalisasi Situs

Personalisasi situs *web* adalah salah satu masalah utama dalam banyak aplikasi berbasis web seperti pemasaran individu, impor elektronik. Rekomendasi kinerja dinamis adalah fitur yang sangat penting dalam banyak hal aplikasi seperti cross-sale dan penjualan di e-commerce. *Web usage mining* adalah penelitian dasar daerah, yang bisa menjadi pendekatan luar biasa untuk ini jenis masalah [49] [50]. Rekomendasi yang ada pada sistem tidak menggunakan *data mining* untuk rekomendasi. Personalisasi situs *web* didasarkan pada penggunaan informasi. *Log server web* dapat digunakan untuk mengelompokkan pengguna *web* memiliki minat terkait. Sistem ini berisi dua modul yang disebut *offline* dan *online* [51].

Modul *offline* membuat *cluster* dengan menggunakan informasi *log* dan *modul online* digunakan untuk tautan generasi dinamis. Semua pengguna situs ditugaskan ke satu cluster berdasarkan pola traversalnya saat ini. Tautan yang ditampilkan di *browser* pengguna mana pun adalah sama dengan pengguna lain di kluster serupa [52] [53].

4.3 Mendukung Desain

Kegunaan adalah masalah yang paling penting di desain internet dan pencapaian situs *web*. Konsekuensi yang dibentuk oleh metode *web usage mining* dapat menyajikan strategi untuk meningkatkan fungsi desain web stereogram memperkirakan asosiasi dan efisiensi situs web dari sudut pandang klien [54]. Metode *web usage mining* untuk mengusulkan variasi yang sesuai untuk situs web. Situs *web* adaptif menjadi ciri langkah selanjutnya. Dalam kasus ini, substansi dan pengaturan situs *web* dapat dengan antusias ditata ulang bersama dengan data yang digali dari perilaku klien [55].

E-commerce melalui *web usage*, intelijen e-commerce dari penggunaan *web* data signifikan secara signifikan untuk berbasis web bisnis.

Keuntungan yang efektif dari penggunaan

Metode *web usage mining* dapat memiliki management klien *relationship*. Dalam hal ini, tujuannya adalah pada masalah-masalah eksplisit bisnis seperti: penghapusan pelanggan, ketertarikan pelanggan, lintas penjualan, dan retensi pelanggan [56] [57] [58].

V. TREN MASA DEPAN

Pengguna WWW perlu menggunakan alat yang cerdas untuk menemukan, mengurutkan, dan memfilter data yang tersedia. *Web mining* bertujuan untuk menemukan dan mengambil informasi di *web*. Dengan menggunakan algoritma, kita dapat menemukan pekerjaan pada data *web*. Pengguna membutuhkan beberapa alat untuk menemukan data yang diperlukan pola penggunaan. Ini memberikan perkembangan cepat sistem cerdas dan cerdas yang dapat menggali Informasi. Menganalisis *file log* sangat sulit proses. sangat sulit untuk menemukan alat yang tepat. Banyak sekali alat tersedia tetapi pengguna tidak suka dan tidak dipertimbangkan sebagai proses yang lambat, kaku, mahal dan terasa sangat sulit mempertahankan. Tetapi beberapa alat digunakan untuk *data mining* sistem untuk membantu menganalisis log, teknik yang ada memiliki banyak hal, apalagi ruang penyimpanan yang besar juga banyak biaya I / O, Sebagian besar alat yang tersedia menyediakan data yang jelas dan statistik. Misinya menjadi lebih menantang saat web tumbuh. Situs web alat pengumpulan informasi sangat besar dan mudah interaktif. Hasil *data mining* akan digabungkan dengan situs web dinamis untuk memberikan tugas yang terkomputerisasi, ujung ke ujung ke Manajemen hubungan pelanggan dan sistem pemasaran. Alat bantu *mining* di sisi yang berbeda adalah penting untuk pengembangan di masa depan. *Web usage mining* harus menangani integrasi informasi offline dengan alat e-commerce, Manajemen *database* relasional, menyediakan katalog produk dan layanan dengan lainnya aplikasi, Beberapa variabel baru diperlukan untuk menemukan pola alami dan bermakna dan berguna. Alat baru adalah ingin banyak waktu proses untuk periode praktek mining. Tes standar selalu perlu untuk algoritma *mining* untuk mengembangkan kinerja karena efektivitas dapat mengukur alat yang lebih baik untuk data web mining.

Untuk mengembangkan visualisasi adalah penting untuk pengguna dapat memahami data. *Web mining* adalah novel dan area penelitian dan aplikasi yang berkembang cepat.

Dengan penelitian di berbagai urutan yang berbeda seperti perangkat lunak

rekayasa, pemasaran, kecerdasan buatan, *database* akan dapat mengembangkan aplikasi *web mining*.

VI. KESIMPULAN

Aplikasi *data mining* banyak digunakan di berbagai bidang seperti *e-business*, pendidikan, telekomunikasi, keuangan, teknik, bioteknologi, anti-terorisme.

Model *web usage mining*, adalah jenis *mining* ke *server log*. *Web usage mining* bertindak sebagai bagian penting dalam mewujudkan kegunaan desain situs *web*, peningkatan hubungan pelanggan, dan peningkatan kebutuhan sistem presentasi dan sebagainya. *Web usage mining* menghadirkan dukungan untuk *server* personalisasi yang menyediakan, desain situs *web*, dan pengambilan keputusan bisnis lainnya. *Web usage mining* terdiri dari tiga fase, yaitu pra-pemrosesan, penemuan pola, dan analisis pola.

Dalam penelitian diperoleh kesimpulan beberapa masalah dengan *web usage mining* seperti menemukan informasi yang diinginkan, menemukan informasi terkait, mempelajari pengetahuan yang berharga, rekomendasi atau personalisasi data, dan penelitian terbaru di bidang penelitian ini. Dari perbandingan *web usage mining* dengan *data mining* di beberapa faktor. Hasilnya bahwa *web usage mining* sangat penting sekarang menjadi hari untuk dunia *web* dan lebih banyak perbaikan diperlukan di masa depan untuk privasi pengguna.

DAFTAR PUSTAKA

- [1] U. Fayyad, "From data mining to knowledge discovery: An overview," in *Proceeding of ACM Knowledge Data Discovery*, 1994.
- [2] E. Spertus, "Mining structural information on the web. Computer Networks and ISDN Systems.," *Int. J. Comput. Telecommuni- cation Netw.*, vol. 29, pp. 1205–1215, 1997.
- [3] M. Pazzan, "Learning from hotlists and coldlists: Towards a www information ltering and seeking agent.," in *In IEEE 1995 International Conference on Tools with Arti cial Intelligence*, 1995.
- [4] M. Balabanovic, "Learning information retrieval agents: Experiments with automated web browsing," in *In On-line Working Notes of the AAAI Spring Symposium Series on Information Gathering from Distributed, Heterogeneous Environments*, 1995.
- [5] O. Bulent, "A Hybrid Web Link Prediction System," in *Proceedings of 2nd International Conference on Intelligent Knowledge Systems (IKS-2005)*, 2005.
- [6] Y. Ho Cho, "Application of Web usage mining and product taxonomy to collaborative recommendations in e-commerce," *Expert Syst. Appl.*, vol. 26, no. 2, pp. 233–246, 2004.

- [7] Rana, "A Study of Web Usage Mining Research Tools," *Int. J. Adv. Netw. Appl.*, vol. 3, no. 6, p. 1442, 2012.
- [8] Z. O, "Discovering Web Access Patterns and Trends by Applying OLAP and Data Mining Technology on Web Logs," in *Proceeding Advances in Digital Libraries Conference*, 1998, pp. 19–29.
- [9] C. Rana, "A Study of Web Usage Mining Research Tools," *Int. J. Adv. Netw. Appl.*, vol. 3, no. 6, pp. 1422–1429, 2012.
- [10] A. Abraham, "Mining, i-Miner: A Web Usage Systems, Framework Using Hierarchical Intelligent," in *IEEE International Conference on Fuzzy Systems FUZZ-IEEE'03*, 2003, pp. 1129–1134.
- [11] R. Cooley, "WebSIFT: The Web Site Information Filter System," in *Proceedings of Workshop on Web Usage Analysis and User Profiling WEBKDD in conjunction with ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1999.
- [12] J. Srivastava, "Web usage mining: Discovery and applications of usage patterns from web data," *SIGKDD Explor. Newsl.*, vol. 1, no. 2, pp. 12–23, 2000.
- [13] C. Shahabi, "INSITE: A Tool for Real-Time Knowledge Discovery from Users Web Navigation," in *Proceedings of the 26th International Conference on Very Large Databases (VLDB)*, 2000, pp. 635–638.
- [14] A. Buchner, "Navigation Pattern Discovery from Internet Data," in *Proceedings of Web Usage Analysis and User Profiling at the International WEBKDD99 Workshop*, 2000, pp. 74–91.
- [15] S. K. Lee, "A Collaborative Recommender System: Lexicographic," *Consens. Web Usage Min. Approach*, vol. 2, 2012.
- [16] O. Nasraoui, "A web usage mining framework for mining evolving user profiles in dynamic web sites," *J. IEEE Trans. Knowl. Data Eng.*, vol. 20, no. 2, pp. 202–215, 2008.
- [17] L. Chen, "WebMate: A Personal Agent for Browsing and Searching," in *Proceedings of the 2nd International Conference on Autonomous Agents*, 1999, pp. 132–139.
- [18] M. Koutri, "A survey on web usage mining techniques for web-based adaptive hypermedia systems," in *Adaptable and Adaptive Hypermedia Systems Idea*, 2004, pp. 1–23.
- [19] M. Eirinaki, "SEWeP: using site semantics and a taxonomy to enhance the Web personalization process," in *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, 2003, pp. 99–018.
- [20] A. . Buchner, "Navigation Pattern Discovery from Internet Data," in *Proceedings of Web Usage Analysis and Use Profiling at the International WEBKDD99 Workshop*, 2000, pp. 74–91.
- [21] C. . Ling, "Data Mining for Direct Marketing: Problems and Solutions," *International J. Electronic Commer. Bus. Media*, vol. 8, no. 3, pp. 32–35, 1998.
- [22] J. Borges, "Data Mining of User Navigation Patterns," in *Web Usage Analysis and User Profiling*, 2000, pp. 31–39.
- [23] P. Senkul, "Improving Pattern Quality in Web Usage Mining by Using Semantic Information," *Knowl. Inf. Syst.*, vol. 1, no. 6, pp. 400–404, 2011.
- [24] K. . Wu, "SpeedTracer: A Web usage mining and analysis tool," *IBM Syst. J. Internet Comput.*, vol. 37, pp. 89–105, 1998.
- [25] M. Rao, "Understanding User Behavior using Web Usage Mining," *Int. J. Comput. Appl.*, vol. 1, no. 7, pp. 56–61, 2010.
- [26] P. Ujwala, "A Survey on User Future Request Prediction: Web Usage Mining," *Int. J. Emerg. Technol. Adv. Eng.*, vol. 2, no. 3, 2012.
- [27] B. Sawan, "Extraction of Business Rules from Web logs to Improve Web Usage Mining," *Int. J. Emerg. Technol. Adv. Eng.*, vol. 2, no. 8, 2012.
- [28] F. T, "Activity monitoring: Noticing interesting changes in behavior.," in *In Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1999, pp. 53–62.
- [29] NetZero, "Netzero." [Online]. Available: <http://www.netzero.com>.
- [30] Alladvantage, "Advantage." [Online]. Available: <http://www.alladvantage.com>.
- [31] E. Cohen, "Improving end-to-end performance of the web using server volumes and proxy lters.," in *Proceeding. ACM SIGCOMM*, 1998, pp. 241–253.
- [32] J. Srivastava, "Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data," *SIGKDD Explor.*, vol. 1, no. 2, pp. 12–23, 2000.
- [33] S. Minocha, "Integrating Customer Relationship Management Strategies in (B2C) E-Commerce Environments," in *IFIP Conference on Human-Computer Interaction-INTERACT*, 2003.
- [34] S. Minocha, "Integrating Customer Relationship Management Strategies in (B2C) ECommerce Environments," in *IFIP Conference on Human-Computer Interaction-INTERACT*, 2003.
- [35] L. Catledge, "Characterizing browsing behaviors on the world wide web.," *Comput. Networks ISDN Syst.*, vol. 27, no. 6, 1995.
- [36] R. Cooley, "Data preparation for mining world wide web browsing patterns," *Knowl. Inf. Syst.*, vol. 1, no. 1, 1991.
- [37] S. G, *Introduction to Modern Information Retrieval*. New York: McGraw-Hil, 1999.
- [38] F. U, "From data mining to knowledge discovery: An overview," in *Proceeding ACM Knowledge Data Discovery*, 1994.
- [39] A. R, "Fast algorithms for mining association

- rules,” in *Proceeding of the 20th VLDB Conference*, 1994, pp. 487–499.
- [40] D. Gibson, “Inferring web communities from link topology,” in *In Conference on Hypertext and Hypermedia*. ACM, 1998.
- [41] K. Etmnani, “Web usage mining: Discovery of the users navigational patterns using SOM,” in *First International Conference on Networked Digital Technologies*, 2009, pp. 224 – 249.
- [42] P. D., “Web Community Directories: A New Approach to Web Personalization,” in *1st European Web Mining Forum*, 2003, pp. 113–129.
- [43] M. D., *Web Usage Mining and Personalization*. CRC Press, 2004.
- [44] D. L. A., “Reengineering Web applications based on cloned pattern analysis,” in *Proceedings 12th IEEE International Workshop on Program Comprehension*, 2004, pp. 132 – 141.
- [45] J. Wang, “Pushing Frequency Constraint to Utility Mining Model,” in *International Conference Computer Science - ICCS 2007*, 2007, pp. 685–692.
- [46] K. M., “Semantic Analysis of Web Pages Using Web Patterns,” in *International Conference on Web Intelligence*, 2006, pp. 329 – 333.
- [47] Eirinaki, “Web mining for web personalization,” *Knowl. Discov.*, vol. 1, no. 1, pp. 52–56, 2008.
- [48] J. Pei, “Mining Access Patterns Efficiently from Web Logs,” in *Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD’00)*, 2000, pp. 396–407.
- [49] T. S. Heng, “Giving meanings to WWW,” in *ACM SIGM Multimedia*, 2000, pp. 39–47.
- [50] J. Ying, “Workshops IEEE Computer Society,” in *International Conference on Data Mining*, 2009.
- [51] J. Wang, “Pushing frequency constraint to utility Mining Model,” in *ICCS Springer-Verlag Berlin Heidelberg*, 2007, pp. 685–692.
- [52] D. Neelam, “Page Ranking Algorithms: A Survey,” in *IEEE International Advance Computing Conference, Patiala*, 2009, pp. 530–1537.
- [53] G. R., “Topological Frequency Utility Mining Model,” in *Springer International Conference, SocPros*, 2011, pp. 505–50.
- [54] L. Chu-Hui, “A novel prediction model based on hierarchical characteristic of web site,” *Expert Syst. Appl.*, p. 38, 2011.
- [55] S. V., “Improved User Navigation Pattern Prediction Technique From Web Log Data,” in *Procedia Engineering*, 2012, p. 30.
- [56] A. A., “A New Web Usage Mining Approach for Next Page Access Prediction,” *Int. J. Comput. Appl.*, vol. 8, no. 11, 2010.
- [57] J. Mehrdad, “WebPUM: A Web-based recommendation system to predict user future movements,” in *Expert Systems with Applications*, 2007, p. 37.
- [58] Trilok Nath Pandey, “Merging Data Mining Techniques for Web Page Access Prediction: Integrating Markov Model with Clustering,” *IJCSI Int. J. Comput. Sci. Issues*, vol. 9, no. 6, 2012.