

PEMANFAATAN *WEB USAGE MINING* DALAM PENGEMBANGAN APLIKASI E-COMMERCE

Kartina Diah KW¹

¹⁾ Teknik Informatika Politeknik Caltex Riau
Jl Umban Sari No.1, Pekanbaru, Riau 28265
Email : diah@pcr.ac.id¹⁾

Abstrak

Proses bisnis pada mulanya diperoleh dari tahapan analisis dengan melakukan requirement elicitation pada tahapan pengembangan aplikasi. Kesulitan melakukan requirement elicitation menjadi tantangan pada saat melakukan pengembangan aplikasi e-commerce, sedangkan proses bisnis e-commerce menjadi semakin kompleks untuk menyesuaikan dengan kebutuhan pelanggan yang semakin kompleks. Hal ini mengakibatkan pekerjaan analisis dalam SDLC membutuhkan solusi yang tepat agar proses bisnis dapat dihasilkan dengan sebisa mungkin meminimalisasi tahapan analisis namun tetap merangkum seluruh kebutuhan pengguna aplikasi e-commerce.

Makalah ini mengusulkan alternatif solusi untuk menghasilkan sebuah model proses bisnis aplikasi e-commerce dari pola akses pengunjung aplikasi e-commerce menggunakan teknik web usage mining yang digunakan untuk mengekstraksi pola yang berasal dari usage data yaitu data log. Teknik yang digunakan untuk melakukan analisa terhadap data log adalah frequent sequence mining. Pola yang dihasilkan dari frequent sequence menunjukkan pola urutan aktivitas pelanggan web e-commerce ketika melakukan aktivitas bisnis berdasarkan waktu kunjungan yang kemudian dimodelkan menjadi sebuah model proses bisnis e-commerce. Model proses yang dihasilkan menunjukkan bahwa tahap analisa dalam SDLC dapat diganti dengan WUM sehingga pengembangan aplikasi e-commerce dapat dilakukan tanpa melakukan requirement elicitation.

Kata kunci: *analisis, requirement elicitation, model proses, web usage mining*

1. Pendahuluan

1.1 Latar Belakang

E-commerce adalah salah satu aplikasi web modern yang dapat melakukan proses pembelian, penjualan, melakukan transfer atau pertukaran barang dan jasa melalui internet, tidak membutuhkan kehadiran fisik dari pelaku tanpa terbatas jarak dan waktu [1], [2].

Desain sebuah aplikasi e-commerce sangat tergantung dari proses bisnis yang akan dijalankannya. Proses bisnis pada mulanya diperoleh dari tahapan analisis dengan

melakukan *requirement elicitation* pada tahapan pengembangan aplikasi, yaitu proses mengumpulkan dan memahami, menemukan, menggali, mempelajari kebutuhan pelanggan, pengguna dan stakeholder sehingga aplikasi yang dikembangkan dapat mengatasi masalah dan memenuhi kebutuhan pelanggan [3]. Kesulitan melakukan *requirement elicitation* menjadi tantangan pada saat melakukan pengembangan aplikasi e-commerce, sedangkan proses bisnis e-commerce menjadi semakin kompleks untuk menyesuaikan dengan kebutuhan pelanggan yang semakin kompleks. Hal ini mengakibatkan pekerjaan analisis dalam SDLC membutuhkan solusi yang tepat agar proses bisnis dapat dihasilkan dengan sebisa mungkin meminimalisasi tahapan analisis namun tetap merangkum seluruh kebutuhan pengguna aplikasi e-commerce.

Dalam memberikan alternatif solusi untuk menggantikan pekerjaan *requirement elicitation* pada tahap analisis dalam SDLC e-commerce dirumuskan permasalahan yang akan diselesaikan sebagai berikut:

1. Bagaimana menghasilkan sebuah model proses dari data log web?
2. Apakah teknik *frequent sequence* dapat digunakan untuk menghasilkan pola akses yang sesuai dengan model proses e-commerce?

Sedangkan tujuan dari makalah ini adalah memberikan alternative solusi dalam pengembangan aplikasi e-commerce pada tahap analisis. Dari alternatif yang disampaikan diharapkan pengembang aplikasi e-commerce dapat menggunakan model proses yang dihasilkan sehingga tahap analisis dalam SDLC dapat tergantikan.

1.2 Tinjauan Pustaka

Proses Bisnis

Seperti halnya dalam aktivitas bisnis konvensional, sistem E-Commerce juga melalui tahapan-tahapan aktivitas tertentu yang biasa diistilahkan dengan proses bisnis. Bedanya, proses bisnis pada aplikasi e-commerce tidak dapat dilepaskan dari urutan navigasi pengguna web ketika menggunakan aplikasi e-commerce. Urutan navigasi digambarkan sebagai urutan akses terhadap halaman web mulai dari awal hingga pengguna meninggalkan sebuah situs web. Namun sebuah proses

bisnis aplikasi e-commerce memiliki karakteristik yang membedakannya dari navigasi, yaitu[4]: pertama, sebuah proses mengarahkan pengguna dalam menjalankan sebuah aktivitas, sebuah proses mendefinisikan sekumpulan aktivitas yang akan dijalankan dan kontrol yang diperlukan selama aktivitas tersebut. Kedua, sebuah proses memiliki pola urutan aktivitas tertentu sehingga dapat menentukan apakah navigasi tertentu yang dilakukan oleh pengguna dapat dianggap sebagai proses atau bukan, misalnya ketika pengguna mengklik tombol *browse* untuk mengambil file atau data tertentu maka aktivitas tersebut tidak dianggap sebagai sebuah urutan proses baru tetapi merupakan satu aktivitas dari aktivitas sebelumnya.

Web Usage Mining

Berdasarkan jenis data yang diekstrak, *web mining* dapat dibagi menjadi tiga kategori, yaitu [5]–[10] :

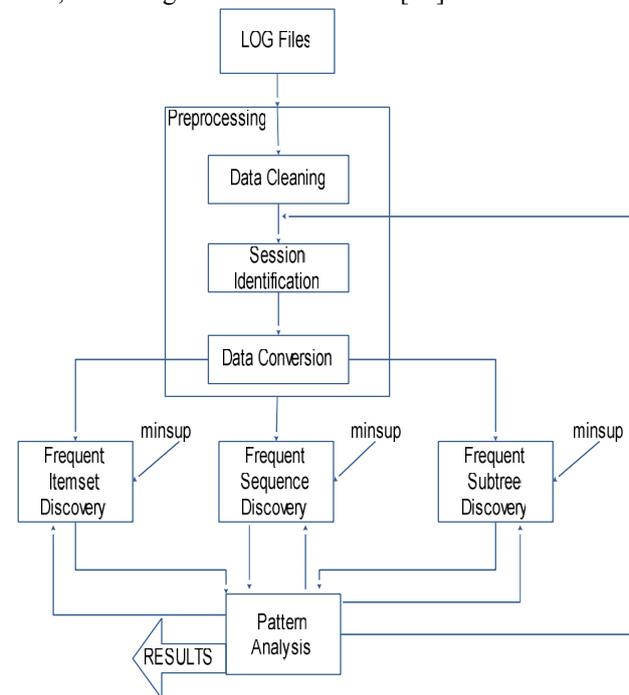
1. *Web Content Mining (WCM)*; teknik *data mining* untuk menghasilkan informasi tentang *content* sebuah dokumen web. Dokumen web tersebut dapat berupa teks, gambar, audio, video, *structured records* seperti *list* dan *table*.
2. *Web Structure Mining (WSM)*; teknik data mining untuk menghasilkan informasi tentang struktur web. Terbagi menjadi 2, yaitu *hyperlinks*, *document structure*.
3. *Web Usage Mining (WUM)*; teknik data mining yang digunakan untuk menemukan pola penggunaan web dari *web usage data*, dengan tujuan untuk memahami dan memberikan layanan yang lebih baik pada aplikasi berbasis web (Srivastava,dkk., 2000). Pola yang dihasilkan berasal dari klasifikasi *usage data* sebagai berikut, *web server data*, *application server data*, *application level data*.

Langkah-langkah yang dilakukan dalam WUM terbagi menjadi 3 tahap, yaitu [8], [12], [13]:

1. *Pre-processing*. Data yang tersedia memiliki kecenderungan mengandung *noise*, tidak lengkap dan tidak konsisten. Pada tahap ini, data akan diolah untuk disesuaikan dengan kebutuhan pada fase berikutnya. Tahap ini meliputi pembersihan data, integrasi data, transformasi data dan reduksi data.
2. *Pattern discovery*. Beberapa metode dan algoritma seperti *statistic*, *data mining*, *machine learning*, dan *pattern recognition* dapat diterapkan untuk mendapatkan pola pada tahap ini.
3. *Pattern Analysis*. Pola yang telah ditemukan selanjutnya dianalisa dan ditampilkan dengan visualisasi dan interpretasi yang lebih difahami oleh pengguna.

Pre-Processing Data

Secara umum tujuan dari WUM adalah untuk memperoleh hasil analisa yang dapat digunakan untuk tahapan desain seperti desain situs web, desain server web, dan navigasi melalui situs web[14].



Gambar 1. Process of Web Usage Mining (Iváncsy & Vajk, 2006)

Log file dari server yang berbeda akan menyimpan informasi yang berbeda-beda. Secara umum data log web berisi informasi pengguna dan aktivitas yang dilakukan. Informasi tersebut terdiri dari *User name*, *IP address*, *date*, *time*, *bytes transferred*, *access request*[14]–[16]. Dari sumber yang sama, Priyanka dan Ujwala mendeskripsikan bahwa ada 4 tipe data log, yaitu *access log file*, *error log file*, *agent log file*, *referer log file*. *Access log file* adalah data semua permintaan yang masuk dan informasi tentang *client server*. *Access log* merekam semua permintaan yang diproses oleh server. *Error log file* adalah daftar kesalahan internal yang tercatat oleh server yang terjadi pada saat terjadi kesalahan ketika *request* yang dikirimkan oleh pengguna. *Agent log file* adalah informasi tentang pengguna *browser*, versi *browser* yang digunakan. *Referer log file* menyediakan informasi tentang link dan pengalihan pengunjung ke halaman web[16].

Tidak semua data logs digunakan pada tahap mining data karena beberapa data tidak relevan dengan kebutuhan terhadap mining yang akan dilakukan, untuk itu diperlukan tahap *preprocessing* data terlebih dahulu. Tahapan *pre-processing* sesuai dengan gambar 1 diatas terdiri dari *data cleaning* berfungsi untuk membersihkan *logs entries* dari data yang tidak dibutuhkan pada proses *mining*. *Session identification* berfungsi untuk mengidentifikasi semua halaman referensi yang

digunakan oleh pengguna untuk disimpan menjadi data dalam *user session*. *Data conversion* berfungsi untuk menyesuaikan format data sesuai dengan kebutuhan pada mining proses.

Pattern Discovery

Proses “*mining*” terjadi pada tahapan ini, meliputi [10], [17]: *statistical analysis, association rules, clustering, classification, frequent pattern, sequential pattern, dependency modeling*.

Frequent Sequence atau yang disebut juga dengan *Sequential Pattern* adalah urutan itemset yang sering terjadi dalam urutan tertentu, semua item dalam itemset yang sama seharusnya berada pada transaksi atau waktu yang sama atau dalam selang waktu yang sama [18]. *Sequential pattern* menemukan pola perilaku navigasi pengguna di halaman web [15]. *Sequential pattern* bertujuan untuk menemukan pola antar-sesi. Pola sekuensial juga mencakup beberapa jenis lain dari analisis sementara seperti analisis tren, mengubah deteksi titik, atau analisis kesamaan.

Generalized Sequential Pattern Algorithm (GSP) [19], *Sequential Pattern Discovery using Equivalent Classes* (SPADE) [20] dan *PrefixSpan* [21] adalah beberapa algoritma yang dapat digunakan untuk melakukan mining dengan *sequential pattern*.

2. Pembahasan

WUM melakukan mining untuk mengekstraksi pola yang berasal dari *usage data* yaitu log data. Log akses web berisi urutan kejadian atau aktivitas atau *events (items)* dalam satu *session*, dengan informasi mencakup *session identifier* dan informasi akses pengguna. Misalkan, dalam data log web terdiri dari informasi beberapa halaman web yang dikunjungi oleh pengunjung dalam 1 *session*. Sebagai contoh, format informasi dari sebuah akses log adalah <SessionId, PageId>, dan data log akses web dari beberapa session sesuai format adalah sebagai berikut:

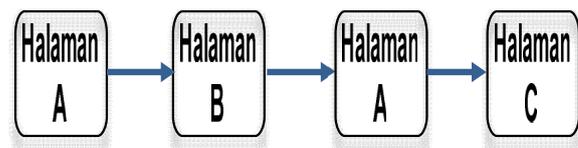
<100,a><100,b><200,e><200,a><300,b><200,e><100,d><200,b><400,a><400,f><100,a><400,b><300,a><100,c><200,c><400,a><200,a><300,b><200,c><300,f><400,c><400,f><400,c><300,a><300,e><300,c>

Selanjutnya data log tersebut dikelompokkan berdasarkan *session*-nya seperti yang terlihat pada tabel 1. Urutan log akses dalam database transaksi pada pada masing-masing baris terdiri dari ID Transaksi (TID) yang berasal dari *session* dan urutan akses (*access sequence*) yang berasal dari *event*. Misalkan untuk *session* 100, *event* yang dilakukan terdiri dari a,b,d,a,dan c.

Tabel 1 Contoh Sequence Database

TID	Web access sequences
100	abdac
200	eaebcac
300	babfaec
400	afbacfc

Dari tabel 1 diatas dapat diperoleh pola urutan akses yang paling sering dilakukan oleh pengunjung. Sebagai contoh, dari data diatas diperoleh bahwa halaman yang paling sering dikunjungi oleh pengunjung adalah sebagai berikut: halaman a = 4, b = 4, c = 4, d = 1, e = 2, f = 2. Sehingga diketahui bahwa setiap pengunjung pasti mengunjungi halaman a, b, c sedangkan halaman d, e, f ada yang mengunjungi ada yang tidak. Selanjutnya dapat juga diketahui pola urutan halaman yang paling sering dikunjungi oleh pengunjung yaitu a,b,a,c. Artinya setiap pengunjung akan mengunjungi halaman web dengan urutan halaman a diikuti halaman b kemudian kembali ke halaman a lalu ke halaman c. Urutan halaman yang dikunjungi tersebut jika digambarkan dalam sebuah graph maka akan terlihat sebagai berikut:



Gambar 1. Alur Kunjungan Halaman Web Dengan WUM

Notasi  menggambarkan halaman web, sedangkan  menunjukkan urutan halaman yang diakses. Analisa ini menjawab permasalahan yang dirumuskan pada poin 1.

Tabel 2 adalah contoh data log yang digunakan pada makalah ini yang diunduh dari <http://lisp.vse.cz/challenge/CURRENT/>. Data log ini adalah data *click stream* yang merekam aktivitas pengguna e-commerce dari 7 aplikasi e-commerce.

Tabel 2 Contoh Data Log Web

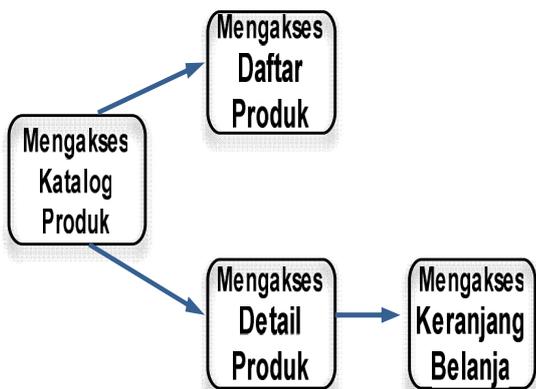
Shop ID	Unixtime()	IP Address	Session	Visited Page
11	1074628802	212.20.106.210	fadb9b666eab173cfc0900148ebd21cb	/
14	1074628802	217.11.233.2	8a8ad32c814ea45c41f1709df7967a19	/ct/?c=164
11	1074628802	82.34.184.151	be5d5d6dc24bf3e54530d7dc8945ff25	/znacka/?&pozice=100&c=102
12	1074628802	80.188.4.42	7a7f25ac8ed0f4e64a1c524803c526e8	/dt/?c=9674&komentare=1
14	1074628803	195.250.148.121	ed03ab09cc0ac7d67a5735599bc6b88e	/dt/?c=7308
11	1074628805	212.20.99.79	2cbb61aefadb5fd25f6d7c9e7a596ce6	/dt/?c=927
10	1074628805	212.111.9.109	344ce202b8094d3aef2fff84cf5da31	/znacka/?&pozice=20&c=2
15	1074628806	194.108.242.78	2feff86721f48bea10510e1594e5ebca	/dt/?c=5231
14	1074628808	213.220.194.175	680dd7669205aed163dbb06bfd8ecabd	/findp/preber3.php
14	1074628808	194.228.166.69	a955d38c204e572ab635418178806168	/ls/preber.php
10	1074628808	62.177.77.138	4e38b17a019a399d018319e337c395ad	/dt/prisl.php?id=7578

Sequential Pattern pada dasarnya tersusun dari himpunan frequent itemset yang merepresentasikan pola berdasarkan urutan waktu terjadinya peristiwa, disebut juga dengan istilah frequent sequence.

Maka pola yang dihasilkan dari analisa yang telah dilakukan dihubungkan dengan konsep aktivitas pengunjung web dalam sebuah session seperti yang telah dijelaskan pada sub bab 3.2.1.3, maka notasi



menggambarkan aktivitas yang dilakukan pada halaman web, sedangkan → tetap menunjukkan urutan halaman yang diakses. Sehingga apabila halaman-halaman web tersebut diumpamakan sebagai halaman-halaman sebuah web e-commerce yang terdiri dari Katalog Produk, Daftar Produk, Detail Produk, dan Keranjang Belanja maka alur prose tersebut akan menjadi seperti gambar dibawah ini.

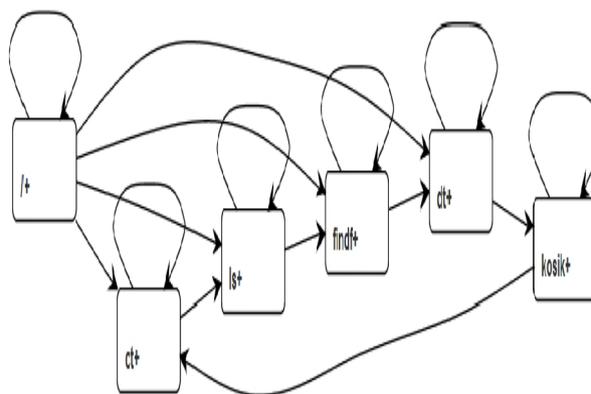


Gambar 2. Pola Akses Halaman Web E-Commerce Dengan Frequent Sequence

Analisa ini menjawab permasalahan yang dirumuskan pada poin 2.

Selanjutnya data yang digunakan diimplementasikan ke dalam algoritma frequent sequence yang digunakan pada makalah ini. Makalah ini menggunakan perangkat ProM dalam mengimplementasikan algoritma mining menggunakan heuristics miner.

Gambar 3 adalah hasil implementasi yang menghasilkan sebuah model proses e-commerce.

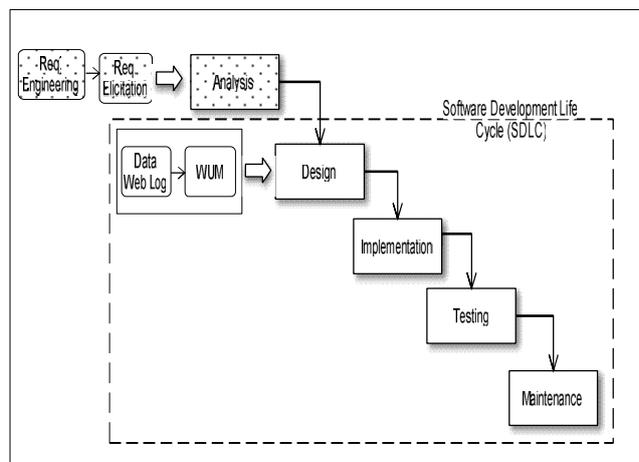


Gambar 1. Model Proses Hasil WUM

Keterangan :

- /+ : Halaman Home
- ct+ : Halaman Kategori Produk
- ls+ : Halaman List Produk
- dt+ : Halaman Detil Produk
- find+ : Halaman Pencarian Produk dan Aksesories
- kosik+ : Halaman Keranjang Belanja

Berdasarkan hasil analisa dan implementasi yang telah dilakukan maka karakteristik pengembangan e-commerce yang diusulkan adalah seperti terlihat pada gambar 4.



Gambar 4. Alur Pengembangan Aplikasi E-Commerce

Gambar diatas adalah gambar tahapan dalam SDLC. Tahapan yang berada dalam garis putus-putus menunjukkan karakteristik pengembangan aplikasi e-commerce yang akan dilakukan. Bagian ini tidak melakukan tahapan analisis yang sebelumnya dilakukan melalui *requirement engineering* dan *requirement elicitation*. Tahapan analisis dilakukan dengan proses menemukan pola kunjungan pengunjung web e-commerce menggunakan teknologi WUM. Dari pola kunjungan yang dihasilkan ternyata memiliki kesesuaian karakteristik dengan sebuah model proses bisnis e-commerce, sehingga pola tersebut akan menjadi model proses e-commerce yang selanjutnya digunakan pada tahap desain untuk mendefinisikan konstruksi struktur halaman web sesuai dengan model proses yang dihasilkan.

Dengan diusulkannya sebuah alternatif solusi dalam pengembangan aplikasi e-commerce maka tujuan dari makalah ini telah tercapai.

3. Kesimpulan

Kesimpulan yang dapat diperoleh dari makalah ini, yaitu :

1. WUM dapat dimanfaatkan untuk menghasilkan model proses bisnis e-commerce dari data log web menggunakan *frequent sequence mining*.
2. Model proses bisnis yang dihasilkan dapat digunakan dalam pengembangan aplikasi e-commerce untuk mengganti tahapan analisis yang sebelumnya dilakukan melalui *requirement elicitation*
3. Dengan menggunakan model proses bisnis menggunakan WUM terdapat beberapa konsekuensi sebagai berikut:
 - a. Mempunyai ketergantungan terhadap data log. Sehingga model proses yang dihasilkan merupakan model proses yang berasal dari aktivitas real pengunjung web e-commerce.
 - b. Model proses yang dihasilkan hanya menggambarkan urutan proses yang dilakukan oleh pengunjung web, konten maupun fungsional detail seperti hapus, tambah, ubah dan sebagainya tidak dapat digambarkan secara spesifik.

Fokus makalah ini hanya untuk membuktikan bahwa WUM dapat digunakan untuk menghasilkan model proses yang digunakan pada tahap pengembangan aplikasi e-commerce, maka saran yang dapat diberikan terkait makalah ini, yaitu:

1. Mencari nilai minimum support yang lebih tepat digunakan untuk mendapatkan sebuah pola kunjungan pengunjung web e-commerce yang lebih ideal sehingga model proses yang dihasilkan juga lebih ideal bagi sebuah web e-commerce.
2. Melakukan pengembangan aplikasi e-commerce dengan melanjutkan ketahap desain hingga pengujian

untuk mengetahui seberapa baik model proses yang dihasilkan dapat menghasilkan aplikasi e-commerce yang efektif dan merangkum kebutuhan pengguna e-commerce.

Daftar Pustaka

- [1] J. Lu and S. S. Gokhale, "Resource Provisioning in an E-commerce Application *," *10th IEEE Conf. E-Commerce Technol. Fifth IEEE Conf. Enterp. Comput. E-Commerce E-Services*, pp. 209–214, 2008.
- [2] N. H. Rawi, M. A. Bakar, R. Bahari, and A. M. Zin, "Development Environment for Layout Design of e-Commerce Applications Using Block-Based Approach," *Int. Conf. Electr. Eng. Informatics*, no. July, 2011.
- [3] A. M. Hickey and A. M. Davis, "Elicitation technique selection: how do experts do it?," *J. Light. Technol.*, pp. 169–178, 2003.
- [4] H. A. Schmid and G. Rossi, "Modeling and Designing Processes in E-Commerce," *IEEE Internet Comput.*, no. February, pp. 19–27, 2004.
- [5] R. Iváncsy and I. Vajk, "Frequent Pattern Mining in Web Log Data," *Acta Polytech. Hungarica*, vol. 3, no. 1, pp. 77–90, 2006.
- [6] N. Kaur, "EXPLORATION OF WEBMINER SYSTEM," *Int. J. Res. IT Manag.*, vol. 2, no. 2, pp. 239–248, 2012.
- [7] R. Kosala, B.-Heverlee, and H. Blockeel, "Web Mining Research: A Survey," *ACM SIGKDD Explor. Newsl.*, vol. 2, no. 1, 2000.
- [8] M. Gomes, "Web Structure Mining: An Introduction," *IEEE Int. Conf. Inf. Acquis.*, pp. 590–595, 2005.
- [9] D. M. Rathod, "A Review On Web Mining," *Int. J. Eng. Res. Technol.*, vol. Vol. 1, no. 2, April 2012, 2012.
- [10] Y. Wang, "Web Mining and Knowledge Discovery of Usage Patterns," *CS748T Proj. (Part I)*, no. Part I, pp. 1–25, 2000.
- [11] J. Srivastava, R. Cooley, M. Deshpande, and P. Tan, "Web usage mining: discovery and applications of usage patterns from Web data," *ACM SIGKDD Explorations Newsletter*, pp. 12–23, 2000.
- [12] R. Cooley, B. Mobasher, and J. Srivastava, "Data Preparation for Mining World Wide Web Browsing Patterns," *Knowl. Inf. Syst.*, vol. 1, no. 1, pp. 5–32, Jul. 2013.
- [13] M. Han, J., Kamber, *Data mining: concept and techniques*. Morgan Kaufmann Publisher, 2000.
- [14] R. Cooley, B. Mobasher, and J. Srivastava, "Data Preparation for Mining World Wide Web Browsing Patterns," *Knowl. Inf. Syst.*, vol. 1, no. 1, pp. 5–32, Jul. 2013.
- [15] L. K. J. Grace and D. Nagamalai, "Analysis Of Web Logs And Web User In Web Mining," *arXiv Prepr. arXiv1101.5668*, vol. 3, no. 1, pp. 99–110, 2011.
- [16] P. Patil and U. Patil, "Preprocessing of web server log file for web mining," *World J. Sci. Technol.*, vol. 2, no. 3, pp. 14–18, 2012.
- [17] J. Cooley, Robert., Srivastava, "Web Usage Mining: Discovery and Application of interesting patterns from web data," University of Minnesota, Minnesota, 2000.
- [18] Q. Zhao, "Sequential Pattern Mining : A Survey," *ITechnical Rep. CAIS Nayang Technol. Univ. Singapore*, no. 2003118, 2003.
- [19] R. Agrawal and R. Srikant, "Mining sequential patterns," *Proc. Elev. Int. Conf. Data Eng.*, pp. 3–14, 1995.
- [20] M. J. Zaki, "SPADE : An Efficient Algorithm for Mining Frequent Sequences," *Mach. Learn.*, vol. 42.1–2, pp. 31–60, 2001.
- [21] J. Pei, J. Han, B. M. Helen, C. Science, C. V. A. E-mail, Q. Chen, and U. Dayal, "PrefixSpan : Mining Sequential Patterns Efficiently by Prefix-Projected Pattern Growth," *EEE 29th Int. Conf. Data Eng.*, 2013.

Biodata Penulis

Kartina Diah Kesuma Wardhani, memperoleh gelar Sarjana Teknik (S.T), Jurusan Teknik Informatika Universitas Islam Indonesia Yogyakarta, lulus tahun 2006. Saat ini menjadi Dosen di Politeknik Caltex Riau, Pekanbaru – Riau.

PENERAPAN TEKNIK DATA MINING UNTUK MENGELOMPOKKAN E-MAIL

Ratih Puspasari¹⁾

¹⁾ *Manajemen Informatika Universitas Potensi Utama
Jl .K.L. Yos Sudarso Km. 6,5 No. 3A Tanjung Mulia Medan
Email : puspalaratih21@yahoo.com¹⁾*

ABSTRAK

Penelitian ini dilakukan di Universitas Potensi Utama Medan dengan menitik beratkan kepada proses pengelompokan *email* dengan menggunakan teknik *Artificial Intelligent Rough Set* yang bertujuan untuk pengambilan keputusan apakah *email* yang ada di STMIK Potensi Utama Medan aman atau tidak dari virus atau ancaman dari luar. Data di ambil dari email di bagian LPPM Universitas Potensi Utama. Data yang telah dikumpulkan kemudian dianalisis dan dipelajari serta dirumuskan sehingga menghasilkan sistem pengambilan keputusan. Sistem pengambilan keputusan tersebut dibantu dengan teknik *Artificial Intelligent Rought Set*. Dari hasil penelitian ini ditemukan bahwa teknik *Artificial Intelligent Rought Set* merupakan sebuah teknik yang dapat diandalkan untuk pengambilan keputusan karena dapat menganalisis data dalam skala besar serta dapat membantu pengambilan keputusan yang tepat waktu.

Kata kunci : *data mining, email, artificial intelligent, roughsSet, pengambilan keputusan.*

1. Pendahuluan

Saat ini email mempunyai peranan yang sangat besar dalam berkomunikasi dengan lebih cepat dan lebih murah dibandingkan dengan metoda komunikasi yang lebih tradisional. Oleh sebab itu email merupakan media komunikasi yang paling banyak digunakan, baik untuk kebutuhan perorangan maupun untuk kebutuhan organisasi, bahkan alamat email sudah merupakan atribut dari identitas seseorang ataupun organisasi.

Dengan relasi yang semakin luas memungkinkan si pemilik alamat email menerima banyak sekali email setiap hari dan terus bertambah tanpa kita sempat membacanya satu persatu. Padahal dari email-email tersebut memungkinkan banyak sekali informasi-informasi berharga yang harus kita ketahui dan banyak pula email-email yang harus hati-hati kita sikapi misalnya email-email yang berisi virus, email yang berisi ancaman atau serangan atau email-email yang berisi rahasia berharga yang tidak bisa diukur nilai kerugiannya dengan uang apabila informasi ini bocor. Bahkan mungkin saja kita pernah menerima email yang tidak perlu dibaca dan hanya memenuhi database email yang kita miliki karena email-email tersebut dikirimkan oleh *spammer*.

Untuk mendapatkan informasi-informasi berharga yang tersembunyi itu diperlukan suatu cara agar dapat menemukan informasi berharga yang tersembunyi dalam tumpukan email yang ada dalam database. Penelitian ini menitik beratkan pada isi dari E-mail yang di terima pada Lembaga Penelitian dan Pengabdian Masyarakat (LPPM) STMIK Potensi Utama. Adapun isi E-mail yang diterima yaitu :

- E-mail yang berisi spam contohnya email yang berupa iklan
- E-mail yang berisi rahasia contohnya email yang berupa pengiriman dan penerimaan bahan makalah.
- E-mail bersifat biasa contohnya email yang berupa informasi dari luar contoh undangan seminar.
- E-mail yang bersifat pribadi

Adapun identifikasi masalah adalah sebagai berikut :

1. Bagaimana menerapkan *Rough Set* untuk mengetahui jumlah pesan-pesan yang dikategorikan sebagai *spam* atau bukan *spam* ?.
2. Bagaimana mengevaluasi kinerja dari *Rough Set* dalam memfilter *email spam*?

Adapun tujuan penelitian ini dilakukan adalah :

1. Untuk mengetahui jumlah e-mail yang masuk berdasarkan kategori *spam* atau bukan *spam*.
2. Sebagai alat penunjang keputusan apakah e-mail LPPM aman atau tidak.

2. Pembahasan

Electronic mail adalah salah satu sarana komunikasi yang cukup handal, perbandingannya dengan mail adalah waktu pengirimannya yang sangat cepat. *Electronic mail* atau disingkat e-mail bukanlah pelayanan "end to end", karena mesin pengirim dan penerima tidak perlu berkomunikasi secara langsung. Proses penyampaian *electronic mail* dapat dianalogikan dengan penyampaian surat oleh Kantor Pos dan Giro. Proses ini disebut "*store and forward*".

Sistem kerja E-mail yang banyak digunakan saat ini banyak menggunakan Unix E-mail model. Model ini mampu untuk mengirim dan menerima E-mail baik di lokal perusahaan (intranet) maupun juga dalam lingkungan global (internet). Unix E-mail Model membagi dalam 3 fungsi yaitu :

1. *Message Transfer Agent (MTA)*
2. *Message Delivery Agent (MDA)*
3. *Message User Agent (MUA)*

MTA, MDA dan MUA beserta E-mail database