

ANALISIS ALGORITMA VMSP PADA MODEL SEQUENTIAL PATTERN DALAM DATA MINING

Ichmi Rianggi Umu Khoirroh¹⁾, Wiwik Suharso²⁾

Teknik Informatika, Teknik, Universitas Muhammadiyah Jember

Jl. Karimata No. 49 Jember, 68121

Telp : (0331) 336728

E-mail: rianggi.ichmi@gmail.com¹⁾, wiwiksuharso@unmuhjember.ac.id²⁾

Abstrak

Algoritma VMSP (*Vertical Of Maximal Sequential Pattern*) adalah algoritma terbaru untuk mendapatkan *maximal sequential pattern* secara *vertical* pada data yang besar. Tujuan dari penelitian ini adalah untuk menganalisis algoritma VMSP agar dapat diketahui tingkat *Accuracy* dan *Precision* dari performa hasil VMSP tersebut. Dataset yang digunakan adalah *web click stream portal berita Kosarak, MNSBC, dan Leviathan*. Dimana dataset yang diinputkan pada sistem dapat menghasilkan *sequential pattern mining* yang maksimal dengan nilai supportnya. Kemudian dilakukan komputasi untuk memperoleh nilai analisis *accuracy* dan *precision* pada performa hasil setiap dataset dengan masing – masing minimal support yang telah ditentukan. Setelah dilakukan komputasi, dataset yang memiliki nilai *accuracy* paling tinggi adalah dataset *Leviathan* dengan minimal support 70%, yaitu 76,94. Pada dataset *Kosarak* memiliki nilai *precision* 100% hingga pada minimal support 60%, *Leviathan* memiliki nilai *precision* 100% hingga pada minimal support 70%, dan *MSNBC* memiliki nilai *precision* 100% hingga pada minimal support 60%.

Kata kunci: VMSP, *Sequential Pattern mining*, *Accuracy*, *Precision*.

Abstract

Algorithm VMSP (*Vertical Of Maximal Sequential Pattern*) is the latest algorithm to obtain *maximal sequential pattern* vertically on large data. The purpose of this study was to analyze the algorithm to be known VMSP *Accuracy* and *precision* level of performance on the VMSP. The dataset used is a news portal *web click stream Kosarak, MNSBC, and Leviathan*. Where dataset is entered on the system can generate *sequential pattern mining* maximum value support. Then do the computation to derive the value of *accuracy* and *precision* analysis on performance of each dataset with each - each predetermined minimal support. After computation, the dataset has the highest value of *accuracy* is dataset *Leviathan* with a minimum support of 70%, ie 76.94. In *Kosarak* dataset has a value of 100% *precision* up on the support of at least 60%, *Leviathan* has *precision* value of 100% up on the support of at least 70%, and *MSNBC* have *precision* value of 100% up on the support of at least 60%.

Keywords: VMSP, *Sequential Pattern mining*, *Accuracy*, *precision*.

1. PENDAHULUAN

Dengan berlimpahnya data yang merupakan akumulasi data yang tersimpan hingga bertahun-tahun pada suatu instansi atau media terutama di internet. Mengakibatkan penumpukan data yang tidak terpakai dalam jumlah besar. Data dalam jumlah besar tersebut dapat dimanfaatkan untuk mendapatkan informasi yang bernilai lebih. Maka, untuk mendapatkan manfaat tersebut dilakukan proses penggalian data (*data mining*)[4]. Dan model dari proses *data mining* sendiri ada banyak. Salah satunya adalah, model yang akan digunakan dalam penelitian ini, yaitu *sequential pattern mining*. Agar dalam melakukan komputasi *sequential pattern mining* dapat dilakukan secara maksimal dan efisien, maka digunakan algoritma VMSP (*Vertical Mining of Maximal Sequential Pattern*) yang terbukti maksimal dan efisien dalam penggunaan ruang memori dan jangka waktu komputasi [3].

Berbeda dengan penelitian sebelumnya oleh Fournier pada tahun 2014 yang mempresentasikan tingkat kecepatan dan keefisienan penggunaan memori dan waktu dalam melakukan komputasi algoritma VMSP, yang menggunakan dataset dan *minimal support* 50%.

Pada penelitian ini mempresentasikan analisis tingkat *precision*, dan *accuracy* dari performa *maximal sequential pattern* setelah dilakukan komputasi algoritma VMSP. Yang menggunakan dataset *web click stream* media Kosarak dengan 10.000 *sequence*, dan Lethiavan dengan 5.834 *sequence* yang juga digunakan pada penelitian sebelumnya, serta menggunakan dataset *web click stream* media MSNBC dengan 31.790 *sequence*. Dimana, dataset ini merupakan *record* data media di internet yang diakses setiap hari, dan dalam satu hari user dapat membaca berbagai jenis berita dari media tersebut. Selain itu, digunakan *minimal support* 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, dan 90%. *Minimal support* tersebut sebagai perwakilan dari berbagai *minimal support* yang diinginkan. Sehingga dapat diketahui hasil tingkat *precision* dan *accuracy maximal sequential pattern* yang ditampilkan setelah dilakukan komputasi algoritma VMSP.

2. TINJAUAN PUSTAKA

2.1 Sequential Pattern Mining

Sequential pattern mining merupakan salah satu model dari data mining, yang digunakan untuk mencari data yang memiliki urutan[2]. Proses *sequential pattern mining* dapat digambarkan sebagai berikut, diberikan sejumlah *sequence* (baris/urutan), setiap *sequence* terdiri atas sejumlah *item* berita, yaitu item berita 1, berita 2, berita 3, berita 4, dan seterusnya, hingga berita n (1,2,3,4,...,n), kemudian diberikan nilai *minimum support* (*minsup*) yaitu batasan nilai minimum dari frekuensi pola yang muncul yang telah ditentukan. *Sequential pattern mining* adalah pencarian semua pola dari item berita yang dibuka oleh user secara berulang, yaitu pola (*pattern*) dari *item* berita yang dibuka oleh user ($\{1\}\{1,2\}, \{2,3\}, \dots, \{n\}$) yang frekuensi kejadiannya lebih besar dari *minimum support* [1].

2.2 Algoritma VMSP

VMSP (*Vertical Mining of Maximal Sequential Pattern*) adalah algoritma pada penggalian data secara vertikal untuk menemukan *sequential pattern mining* yang maksimal[3]. Langkah – langkah algoritma VMSP [3], yang pertama adalah mengambil input sebuah *sequence* yang ada dalam *dataset web click stream* dan batasan *minsup* yang kemudian menjadi *vertical mining sequential pattern*. Pada penelitian ini *vertical mining sequential pattern* ditentukan dengan pola yang telah terbentuk, yaitu pola dari item berita apa saja yang dibuka oleh user dan *user* mana saja yang telah membuka pola dari item berita tersebut dengan *support* \geq batasan *minsup*. Kedua, dilakukan pencarian pola maksimal (*maximal pattern*). *Maximal pattern* adalah ketika salah satu pola dalam baris (*sequential*) yang satu berada pada baris yang lain dan memiliki jumlah pola yang sama.

Max sequential pattern = pola dalam *sequential* awal (A) \subseteq pola dalam *sequential* selanjutnya (B)
(1)

Setelah didapatkan pola maksimal, langkah selanjutnya dilakukan strategi – strategi algoritma VMSP. Ada 3 strategi pada algoritma VMSP yaitu:

- 1) *Efficient Filtering of Non-maximal pattern* (EFN), yaitu memfilter *sequential pattern* yang tidak maksimal, dengan hanya menampilkan *sequential pattern* yang maksimal (persamaan 1 dan 2).
- 2) *Forward-Maximal Extension checking* (FME). Pencarian prosedur menemukan *pattern* dengan menumbuhkan sebuah *pattern* dengan menambahkan satu *item* dalam satu waktu dengan *s-extension* atau *i-extension*. Dimana *s-extension* adalah ketika dalam sebuah *sequential pattern* memiliki beberapa pola yang terbentuk, seperti ($\{1\}\{1,2\}, \{2,3\}, \dots, \{n\}$). Sedangkan *i-extension* adalah ketika dalam sebuah pola memiliki beberapa subpola (*subpattern*), seperti ($\{1,2,3, \dots, n\}$).
- 3) *Candidate Pruning by Co – Occurrence map* (CPC). Strategi ini bertujuan dalam pemangkasan ruang pencarian dari pola. Yaitu dengan pemangkasan *s-extension* dan *i-extension*, dimana ketika salah satu *item* tidak dapat membentuk pola yang maksimal.

2.3 Analisis Accuracy dan Precision

Analisis kinerja algoritma yang digunakan dalam penelitian ini adalah sebagai berikut:

- 1) *Precision* adalah proporsi dari hasil komputasi relevan dari algoritma yang ditampilkan (relevan dan ditampilkan) terhadap semua hasil komputasi relevan yang ditampilkan algoritma (relevan dan ditampilkan) dan hasil komputasi relevan yang tidak ditampilkan algoritma (relevan dan tidak ditampilkan)[5].
- 2) *Accuracy* adalah proporsi dari hasil komputasi relevan dari algoritma yang ditampilkan (relevan dan ditampilkan) terhadap semua kondisi hasil komputasi yang relevan dan tidak relevan, serta yang ditampilkan dan tidak ditampilkan[6].

$$\text{Precision} = \frac{TP}{TP + FP} \times 100\% \quad (2)$$

$$\text{Accuracy} = \frac{TP}{TP + TN + FP + FN} \times 100\% \quad (3)$$

TP (relevan dan ditampilkan), adalah pola dengan nilai *support* yang dimilikinya. Dan pola yang ditampilkan ≥ 1 . Maka, untuk mendapatkan nilai hasil komputasi relevan yang ditampilkan algoritma dari *dataset* yang diinputkan adalah jumlah keseluruhan nilai *support* dari pola yang ditampilkan, dibagi dengan jumlah *sequence* pola yang ditampilkan. TN (relevan dan tidak ditampilkan), adalah 0. Karena semua hasil komputasi relevan ditampilkan. FP (tidak relevan dan ditampilkan), adalah 0. Karena semua hasil komputasi yang ditampilkan adalah yang relevan. FN (tidak relevan dan tidak ditampilkan), adalah jumlah keseluruhan *sequential pattern* yang dimiliki *dataset* yang telah diinputkan, dikurangi dengan jumlah keseluruhan nilai *support* dari pola yang ditampilkan dibagi dengan jumlah *sequence* pola yang ditampilkan.

3. METODOLOGI PENELITIAN

3.1 Persiapan Penelitian

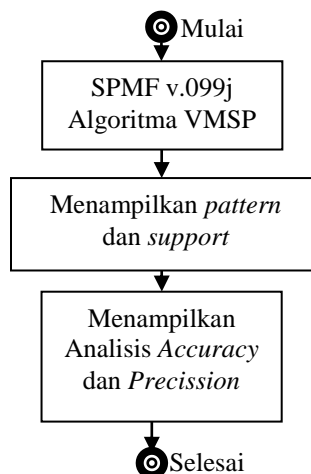
Pada persiapan penelitian, dilakukan studi pustaka untuk mendapatkan dasar – dasar referensi yang berkaitan dengan teori algoritma VMSP. Dan juga dilakukan pengumpulan data. Data yang digunakan adalah data record dari *web click stream* portal berita Kosarak dengan 10. 000 *sequence*, dan Lethiavan dengan 5. 834 *sequence* yang juga digunakan pada penelitian sebelumnya, serta menggunakan dataset *web click stream* media MSNBC dengan 31.790 *sequence*. Semua *dataset* merupakan *record* data media di internet yang diakses setiap hari, dan *user* membaca berbagai jenis berita dari media tersebut.

3.2 Analisis Sistem

Pada tahap ini, akan dilakukan analisis kebutuhan sistem. Dimana sistem dapat menerima dokumen digital yang bertipe data .txt yang berisi *sequential pattern* dari setiap *dataset*. System yang digunakan adalah SPMF v.099j yang merupakan *open source data mining library* menggunakan bahasa pemrograman java[8]. Pada penelitian ini SPMF v.099j dikembangkan dengan menambahkan tampilan *form* dari analisis *precision* dan *accuracy*, agar dapat diketahui tingkat *accuracy* dan *precision* performa hasil *maximal sequential pattern* dengan algoritma VMSP.

3.3 Rancangan Sistem

Untuk mempermudah komputasi dengan *dataset* yang besar, maka digunakan sistem. Rancangan system yang digunakan adalah sebagai berikut:



Gambar 1. Rancangan system

Pertama digunakan SPMF v.099 Algoritma VMSP, untuk mendapatkan hasil dari komputasi VMSP. Selanjutnya sistem menampilkan hasil yang berisi *maximal sequential pattern (pattern)* beserta nilai supportnya (#SUP). Dimana nilai *support* adalah jumlah banyaknya setiap *maximal sequential pattern* dari *dataset*. Setelah sistem menampilkan hasil VMSP. Selanjutnya dapat diketahui analisis kinerja algoritma. Analisis kinerja yang digunakan adalah analisis *precision* dan *accuracy*. Cara mudah membuat layout adalah dengan menggunakan panduan ini secara langsung.

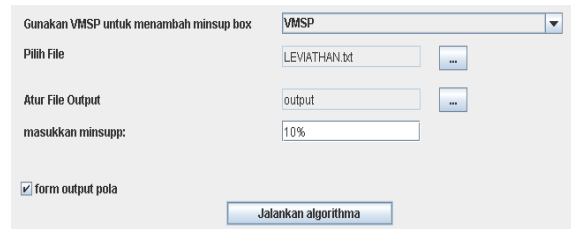
3.4 Pembuatan Sistem

Pembuatan sistem ini menggunakan Netbeans 8.0.2.

4. HASIL DAN PEMBAHASAN

4.1 Hasil

Tampilan SPMF v.099 Algoritma VMSP yang digunakan untuk melakukan komputasi secara otomatis adalah sebagai berikut:



Gambar 2. SPMF v. 099 Algoritma VMSP

Pada halaman awal ini digunakan algoritma VMSP, kemudian pilih *file dataset* yang akan digunakan, setelah itu buat *file output*, dilanjutkan tentukan *minimal support (minsup)*, centang *form output pola* untuk menampilkan hasil output akhir pola, dan klik button *run algorithm* untuk menjalankan system. Maka akan tampil halaman hasil dari pola (*Pattern*) dan nilai *support (#SUP)*.

Hasil dari *dataset Kosarak* yang menampilkan hasil output hanya dengan *minimal support* 10%, 20%, 30%, 40%, 50%, 60%, karena pola yang memiliki nilai *support* diatas 70%, 80%, dan 90% adalah 0 (tidak dimiliki oleh *dataset Kosarak*).

Tabel 1. Hasil *pattern* dan nilai *support* dengan *minsup* 10%, 20%, 30%, 40%, 50% pada *dataset Kosarak*

No	Minsup10%		Minsup20%		Minsup30%		Minsup40%		Minsup50%	
	Pattern	#SUP	Pattern	#SUP	Pattern	#SUP	Pattern	#SUP	Pattern	#SUP
1	{1}{6}	1.272	{11}{6}	3.260	{3}	4.569	{6}	6.058	{6}	6.058
2	{11}{6}{3}	1.421	{6}{3}	2.676	{11}{6}	3.260	{3}	4.569		

Pada table 1, *dataset Kosarak* dengan *minimal support* 10%, pola berita yang sering dikunjungi adalah berita 11, berita 6, dan berita 3 sebanyak 1.421, dan seterusnya hingga *minimal support* 50% pola berita yang sering dikunjungi adalah berita 6 sebanyak 6.058. Pada tabel 2, *dataset Kosarak* dengan *minimal support* 60%, pola berita yang sering dikunjungi adalah berita 6 sebanyak 6.058, dan seterusnya hingga *minimal support* 90% pola berita yang sering dikunjungi adalah berita 0 sebanyak 0.

Tabel 2. Hasil *pattern* dan nilai *support* dengan *minsup* 60%, 70%, 80%, 90% pada portal Kosarak

No	Minsup60%		Minsup70%		Minsup80%		Minsup90%	
	Pattern	#SUP	Pattern	#SUP	Pattern	#SUP	Pattern	#SUP
1	{6}	6.058	0	0	0	0	0	0

Hasil dari *dataset Leviathan* yang menampilkan hasil *output* hanya dengan *minimal support* 10%, 20%, 30%, 40%, 50%, 60%, 70% karena pola yang memiliki nilai *support* diatas 80%, dan 90% adalah 0 (tidak dimiliki oleh *dataset Leviathan*).

Tabel 3. Hasil *pattern* dan nilai *support* dengan *minsup* 10%, 20%, 30%, 40%, 50% pada *dataset Leviathan*

No	Minsup10%		Minsup20%		Minsup30%		Minsup40%		Minsup50%	
	Pattern	#SUP	Pattern	#SUP	Pattern	#SUP	Pattern	#SUP	Pattern	#SUP
1	{347}	587	{290}	1.715	{40}	1.916	{71}	3.383	{71}	3.383
2	{227}	1.001	{124}	1.183			{29}	2.898	{17}	3.026
3	{224}	628	{103}	1.371			{21}	2.467	{14}	3.260
4	{212}	753	{96}	1.538			{17}	3.026		
5	{206}	766	{92}	1.422			{14}	3.260		
6	{197}	685	{52}	1.723						
7	{75}	652	{39}	1.430						
8	{63}	803	{30}	1.582						
9	{54}	862	{20}	1.475						
10	{36}	637	{3}	1.693						

No	Minsup10%		Minsup20%		Minsup30%		Minsup40%		Minsup50%	
	Pattern	#SUP	pattern	#SUP	pattern	#SUP	pattern	#SUP	pattern	#SUP
11	{26}	722								
12	{23}	686								
13	{19}	831								
14	{11}	792								
15	{18}{318}	609								
16	{8}{318}	654								
17	{18}{227}	705								
18	{8}{227}	727								
19	{18}{122}	661								
20	{8}{122}	697								

Pada tabel 3, *dataset* Leviathan dengan *minimal support* 10%, pola berita yang sering dikunjungi adalah berita 8 dan berita 122 sebanyak 697, dan seterusnya hingga *minimal support* 50% pola portal berita yang sering dikunjungi adalah berita 71 sebanyak 3.383. Pada tabel 4, *dataset* Leviathan dengan *minimal support* 60%, pola berita yang sering dikunjungi adalah berita 8 dan berita 18 sebanyak 3.522, dan seterusnya hingga *minimal support* 90% pola berita yang sering dikunjungi adalah berita 0 sebanyak 0.

Tabel 4. Hasil *pattern* dan nilai *support* dengan *minsup* 60%, 70%, 80%, 90% pada *dataset* Leviathan

No	Minsup60%		Minsup70%		Minsup80%		Minsup90%	
	Pattern	#SUP	pattern	#SUP	pattern	#SUP	pattern	#SUP
1	{8}{18}	3.522	{18}	4.326	0	0	0	0
2			{8}	4.651				

Hasil dari *dataset* MSNBC yang menampilkan hasil *output* hanya dengan *minimal support* 10%, 20%, 30%, 40%, 50%, 60%, karena pola yang memiliki nilai *support* diatas 70%, 80%, dan 90% adalah 0 (tidak dimiliki oleh *dataset* MSNBC).

Tabel 5. Hasil *pattern* dan nilai *support* dengan *minsup* 10%, 20%, 30%, 40%, 50% pada *dataset* MSNBC

No	Minsup10%		Minsup20%		Minsup30%		Minsup40%		Minsup50%	
	Pattern	#SUP	Pattern	#SUP	Pattern	#SUP	Pattern	#SUP	Pattern	#SUP
1	{8}	4.942	{10}	9.564	{14}	10.240	{7}	15.198	{2}	20.695
2	{1}{15}	3.227	{1}{10}	7.458	{12}	11.785	{9}	13.969	{1}{2}	16.785
3	{2}{5}	3.479			{11}	10.507	{6}	15.360		
4	{1}{5}	4.162			{10}	9.564				
5					{6}	13.969				
6					{4}	15.360				
7					{3}	9.921				
8					{1}{12}	9.613				
9					{1}{4}	11.142				

Pada tabel 5, *dataset* MSNBC dengan *minimal support* 10%, pola berita yang sering dikunjungi adalah berita 8 sebanyak 4.942, dan seterusnya hingga *minimal support* 50% pola portal berita yang sering dikunjungi adalah berita 2 sebanyak 20.695. Pada tabel 6, *dataset* MSNBC dengan *minimal support* 60%, pola berita yang sering dikunjungi adalah berita 2 sebanyak 20.695, dan seterusnya hingga *minimal support* 90% pola berita yang sering dikunjungi adalah berita 0 sebanyak 0.

Tabel 6. Hasil *pattern* dan nilai *support* dengan *minsup* 60%, 70%, 80%, 90% pada *dataset* MSNBC

No	Minsup60%		Minsup70%		Minsup80%		Minsup90%	
	Pattern	#SUP	pattern	#SUP	pattern	#SUP	pattern	#SUP
1	{2}	20.695	0	0	0	0	0	0

Dari hasil *pattern* dan nilai *support* diatas (tabel 1, tabel 2, tabel 3, tabel 4, tabel 5, dan tabel 6) diperoleh hasil analisis *accuracy* pada masing – masing *dataset*, dan menggunakan *minimal support* 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, dan 90%, dengan menggunakan komputasi pada persamaan 3. Maka dihasilkan nilai prosentase *accuracy* seperti pada tabel 7.

Tabel 7. Tabel analisis *Accuracy*

Minsupp (%)	10	20	30	40	50	60	70	80	90
Kosarak (%)	13,47	29,68	39,15	53,14	60,58	60,58	0	0	0
Leviathan (%)	12,39	25,94	32,84	51,53	55,25	60,37	76,94	0	0
MSNBC (%)	12,43	26,77	35,69	46,69	58,95	65,1	0	0	0

Pada tabel 7, *dataset* Kosarak dengan *minimal support* 10% memiliki tingkat *accuracy* 13,47%, dan seterusnya hingga pada *dataset* MSNBC dengan *minimal support* 90% memiliki tingkat *accuracy* 0%.

Dari hasil *pattern* dan nilai *support* diatas (tabel 1, tabel 2, tabel 3, tabel 4, tabel 5, dan tabel 6) diperoleh hasil analisis *precision* pada masing – masing *dataset*, dan menggunakan *minimal support* 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, dan 90%, dengan menggunakan komputasi pada persamaan 2. Maka dihasilkan nilai prosentase *precision* seperti pada table 8.

Tabel 8. Tabel analisis *precision*

Minsupp (%)	10	20	30	40	50	60	70	80	90
Kosarak (%)	100	100	100	100	100	100	0	0	0
Leviathan (%)	100	100	100	100	100	100	100	0	0
MSNBC (%)	100	100	100	100	100	100	0	0	0

Pada tabel 8, *dataset* Kosarak dengan *minimal support* 10% memiliki tingkat *precision* 100%, dan seterusnya hingga pada *dataset* MSNBC dengan *minimal support* 90% memiliki tingkat *precision* 0%

4.2 Pembahasan

Dari hasil diatas, didapatkan bahwa:

- 1) Prosentase dari analisis *accuracy* dari *dataset web click stream* Kosarak tertinggi adalah dengan *minimal support* 60%, yaitu dengan prosentase *accuracy* 60,58%. Leviathan adalah dengan *minimal support* 70%, yaitu dengan prosentase *accuracy* 76,94%. Dan MSNBC adalah dengan *minimal support* 60%, yaitu dengan prosentase *accuracy* 65, 1%.
- 2) Prosentase tertinggi dari analisis *accuracy* diantara *dataset web click stream* Kosarak, Leviathan, dan MSNBC adalah *dataset* Leviathan dengan *minimal support* 70% yaitu 76,94%.
- 3) *Dataset* Kosarak memiliki nilai *precision* 100% hingga pada *minimal support* 60%, Leviathan memiliki nilai *precision* 100% hingga pada *minimal support* 70%, dan MSNBC memiliki nilai *precision* 100% hingga pada *minimal support* 60%. Hal ini dikarenakan, pada *dataset* kosarak dengan *minimal support* 70%, 80%, dan 90% bernilai 0 (tidak memiliki *output* hasil yang ditampilkan), Leviathan dengan *minimal support* 80% dan 90% bernilai 0 (tidak memiliki *output* hasil yang ditampilkan), dan MNSBC dengan *minimal support* 70%, 80%, dan 90% bernilai 0 (tidak memiliki *output* hasil yang ditampilkan)

5. SIMPULAN DAN SARAN

5.1 Simpulan

Prosentase tertinggi dari analisis *accuracy* dari *dataset web click stream* Kosarak, Leviathan, dan MSNBC adalah *dataset* Leviathan dengan *minimal support* 70% yaitu 76,94%. *Dataset* Kosarak memiliki nilai *precision* 100% hingga pada *minimal support* 60%, Leviathan memiliki nilai *precision* 100% hingga pada *minimal support* 70%, dan MSNBC memiliki nilai *precision* 100% hingga pada *minimal support* 60%.

5.2 Saran

Melihat hasil komputasi analisis *accuracy* dan *precision* pada penelitian ini. Penelitian ini masih dapat dikembangkan lagi dengan menggunakan *dataset* yang lain dan yang lebih bervariasi untuk mendapatkan hasil yang lebih baik.

6. DAFTAR RUJUKAN

- [1] Hani'ah, Mamluatul. Ratnawati, Dian, Eka. dan Suprpto. 2014. Penggalan Pola Sekuensial pada data akses pengguna *website* menggunakan Algoritma PrefixSpan. Malang: Universitas Brawijaya.
- [2] Ayres, Jay. Gehrke, Jhannes. Yiu, Tomi. dan Flannck, Jason. 2002. *Sequential Pattern Mining using A Bitmap Representation*. Cornell University.

- [3] Fournier, Viger, P. Wu – Cheng, W. Gomariz, Antonio, dan Tseng, Vincent S. 2014. *VMSP: Efficient Vertical Mining of Maximal Sequential Patterns*. Canada: Universite de Montreal.
- [4] Han, J. Kamber, M. 2001. *Data Mining: Concepts and Techniques*. California.
- [5] Muningsih, Elly. 2016. *Penentuan Rekomendasi Produk dengan Metode Data Mining Asosiasi Generalized Sequence Pattern (GSP)*. Yogyakarta: AMIK BSI.
- [6] Julastio, Riky. Gunawan. 2015. *Sequential Pattern Mining Dengan SPADE untuk Prediksi Pembelian Spare Part dan Aksesori Komputer pada Kedatangan Kembali Konsumen*. Surabaya: Sekolah Tinggi Teknik Surabaya.
- [7] Sulisty, Wiwin. 2012. *Pemodelan Kesesuaian Dokumen Sekuriti Manajemen Aset Teknologi Informasi Menggunakan Algoritma Extended Weighted – Tree Similarity*. Surabaya.
- [8] Fournier, Viger, P. 2016. *SPMF An Open Source Data Mining Library*. [Online] (Updated 16 Juni 2016). Available at: <http://www.philippe-fournier-viger.com/spmf/>. [Accessed 10 Juli 2016]

Halaman ini sengaja dikosongkan