

**PENGUJIAN ALGORITMA SEQUENTIAL
PATTERN MINING DENGAN LIBRARY SPMF**

Skripsi



oleh
NICO RAYNALDO
71130041

PROGRAM STUDI INFORMATIKA FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS KRISTEN DUTA WACANA
2018

**PENGUJIAN ALGORITMA SEQUENTIAL
PATTERN MINING DENGAN LIBRARY SPMF**

Skripsi



Diajukan kepada Program Studi Informatika Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana
Sebagai Salah Satu Syarat dalam Memperoleh Gelar
Sarjana Komputer

Disusun oleh

NICO RAYNALDO
71130041

**PROGRAM STUDI INFORMATIKA FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS KRISTEN DUTA WACANA 2018**

PERNYATAAN KEASLIAN SKRIPSI

Saya menyatakan dengan sesungguhnya bahwa skripsi dengan judul:

PENGUJIAN ALGORITMA SEQUENTIAL PATTERN MINING (SPAM) DENGAN LIBRARY SPMF

yang saya kerjakan untuk melengkapi sebagian persyaratan menjadi Sarjana Komputer pada pendidikan Sarjana Program Studi Informatika Fakultas Teknologi Informasi Universitas Kristen Duta Wacana, bukan merupakan tiruan atau duplikasi dari skripsi kesarjanaan di lingkungan Universitas Kristen Duta Wacana maupun di Perguruan Tinggi atau instansi manapun, kecuali bagian yang sumber informasinya dicantumkan sebagaimana mestinya.

Jika dikemudian hari didapati bahwa hasil skripsi ini adalah hasil plagiasi atau tiruan dari skripsi lain, saya bersedia dikenai sanksi yakni pencabutan gelar kesarjanaan saya.

Yogyakarta, 15 Oktober 2018



NICO RAYNALDO

71130041

HALAMAN PERSETUJUAN

Judul Skripsi : PENGUJIAN ALGORITMA SEQUENTIAL
PATTERN MINING (SPAM) DENGAN LIBRARY
SPMF

Nama Mahasiswa : NICO RAYNALDO

N I M : 71130041

Matakuliah : Skripsi (Tugas Akhir)

Kode : TIW276


Semester : Genap

Tahun Akademik : 2017/2018

Telah diperiksa dan disetujui di
Yogyakarta,
Pada tanggal 15 Oktober 2018

Dosen Pembimbing I

Dosen Pembimbing II



Budi Susanto, SKom.,M.T.



Gloria Virginia, S.Kom., MAI, Ph.D.

HALAMAN PENGESAHAN

PENGUJIAN ALGORITMA SEQUENTIAL PATTERN MINING (SPAM) DENGAN LIBRARY SPMF

Oleh: NICO RAYNALDO / 71130041

Dipertahankan di depan Dewan Penguji Skripsi
Program Studi Informatika Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana - Yogyakarta
Dan dinyatakan diterima untuk memenuhi salah satu syarat memperoleh gelar
Sarjana Komputer
pada tanggal 26 Juli 2018


Yogyakarta, 15 Oktober 2018
Mengesahkan,

Dewan Penguji:

1. Budi Susanto, SKom., M.T.
2. Gloria Virginia, S.Kom., MAI, Ph.D.
3. R. Gunawan Santosa, Drs. M.Si.
4. Aditya Wikan Mahastama, S.Kom., M.Cs.



Dekan


(Budi Susanto, S.Kom., M.T.)

Ketua Program Studi


(Gloria Virginia, Ph.D.)

INTISARI

PENGUJIAN ALGORITMA SEQUENTIAL PATTERN MINING (SPAM) DENGAN LIBRARY SPMF

Pada zaman yang modern ini, perkembangan dunia bisnis sangatlah pesat. Salah satu bisnis yang kini mulai berkembang pesat yaitu bisnis ritel. Perilaku pembelian konsumen tentunya akan sangat berpengaruh dalam kegiatan bisnis perusahaan. Oleh karena itu, penting bagi perusahaan untuk mengetahui pola pembelian konsumen sehingga perusahaan dapat mengetahui produk-produk apa saja yang perlu dipersiapkan dan diminati oleh konsumen. Namun, menganalisis pola sekuensial pembelian dari data yang besar tidaklah mudah. Kendala yang dialami yaitu seperti lamanya waktu yang dibutuhkan dalam menganalisis pola sekuensial. Karena apabila terlalu lama mungkin hasil analisis pola sekuensial tersebut sudah tidak valid.

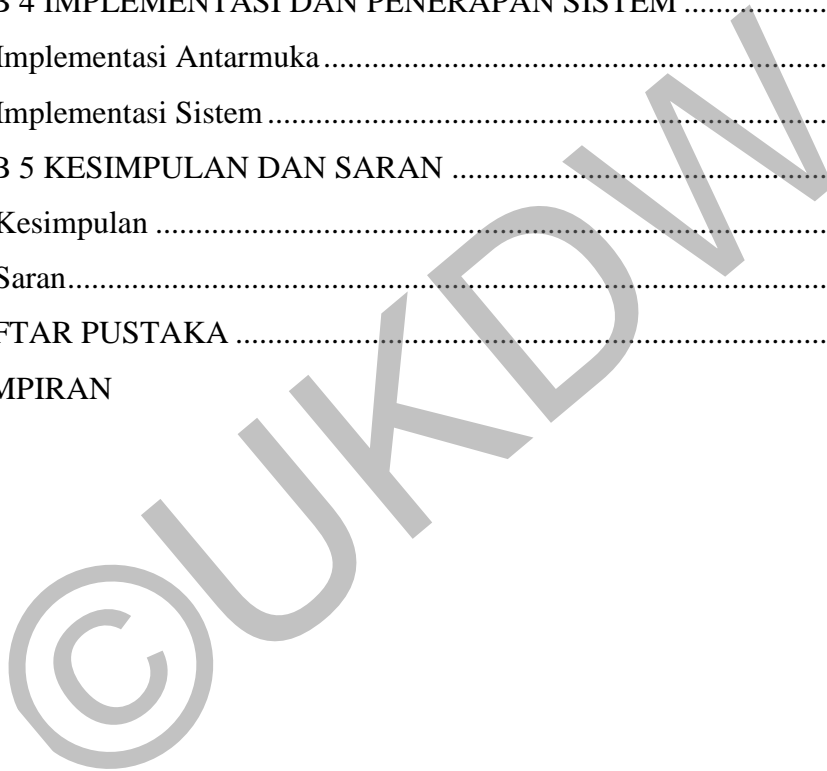
Oleh karena itu, maka untuk menganalisis pola sekuensial tersebut dapat menggunakan algoritma SPAM. Algoritma SPAM dapat mengolah data yang besar secara lebih cepat dan efektif jika dibandingkan beberapa algoritma lain seperti algoritma SPADE dan *Prefix Span*. Algoritma SPAM menggunakan representasi *bitmap* vertikal dalam prosesnya. Algoritma SPAM juga menggunakan metode *pruning* dalam prosesnya untuk menyingkirkan pola-pola yang jarang muncul sehingga meningkatkan efisiensi program.

Dapat disimpulkan bahwa algoritma SPAM cocok digunakan untuk pengolahan data yang besar. Namun karena dalam penggunaannya diperlukan memory yang cukup besar juga, maka untuk mengolah data yang besar disarankan untuk menjalankan algoritma SPAM ini pada perangkat keras yang memiliki tingkat kinerja yang tinggi semisal komputer server.

Daftar Isi

HALAMAN JUDUL.....	i
PERNYATAAN KEASLIAN SKRIPSI.....	iii
HALAMAN PERSETUJUAN.....	iv
HALAMAN PENGESAHAN.....	v
INTISARI.....	vi
DAFTAR ISI.....	vii
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang Masalah.....	1
1.2 Perumusan Masalah	2
1.3 Batasan Masalah.....	2
1.4 Tujuan Penelitian	3
1.5 Metode/Pendekatan.....	3
1.6 Sistematika Penulisan	4
BAB 2 TINJAUAN PUSTAKA	5
2.1 Tinjauan Pustaka	5
2.2 Landasan Teori.....	6
2.2.1 Data Mining	6
2.2.2. SPAM.....	6
2.2.3. Lexicographic Tree Sequence	7
2.2.4 Studi Kasus Algoritma SPAM	8
BAB 3 ANALISIS DAN PERANCANGAN SISTEM	12
3.1 Analisis Kebutuhan	12
3.1.1 Kebutuhan Fungsional	12

3.1.2 Kebutuhan Non Fungsional	12
3.2 Use Case Diagram.....	12
3.3 Alur Kerja Program.....	14
3.4 Rancangan Environment	15
3.5 Rancangan Proses.....	15
3.6 Rancangan Antarmuka	16
3.7 Rancangan Pengujian	17
BAB 4 IMPLEMENTASI DAN PENERAPAN SISTEM	18
4.1 Implementasi Antarmuka	18
4.2 Implementasi Sistem	22
BAB 5 KESIMPULAN DAN SARAN	50
5.1 Kesimpulan	50
5.2 Saran.....	50
DAFTAR PUSTAKA	51
LAMPIRAN	



INTISARI

PENGUJIAN ALGORITMA SEQUENTIAL PATTERN MINING (SPAM) DENGAN LIBRARY SPMF

Pada zaman yang modern ini, perkembangan dunia bisnis sangatlah pesat. Salah satu bisnis yang kini mulai berkembang pesat yaitu bisnis ritel. Perilaku pembelian konsumen tentunya akan sangat berpengaruh dalam kegiatan bisnis perusahaan. Oleh karena itu, penting bagi perusahaan untuk mengetahui pola pembelian konsumen sehingga perusahaan dapat mengetahui produk-produk apa saja yang perlu dipersiapkan dan diminati oleh konsumen. Namun, menganalisis pola sekuensial pembelian dari data yang besar tidaklah mudah. Kendala yang dialami yaitu seperti lamanya waktu yang dibutuhkan dalam menganalisis pola sekuensial. Karena apabila terlalu lama mungkin hasil analisis pola sekuensial tersebut sudah tidak valid.

Oleh karena itu, maka untuk menganalisis pola sekuensial tersebut dapat menggunakan algoritma SPAM. Algoritma SPAM dapat mengolah data yang besar secara lebih cepat dan efektif jika dibandingkan beberapa algoritma lain seperti algoritma SPADE dan *Prefix Span*. Algoritma SPAM menggunakan representasi *bitmap* vertikal dalam prosesnya. Algoritma SPAM juga menggunakan metode *pruning* dalam prosesnya untuk menyingkirkan pola-pola yang jarang muncul sehingga meningkatkan efisiensi program.

Dapat disimpulkan bahwa algoritma SPAM cocok digunakan untuk pengolahan data yang besar. Namun karena dalam penggunaannya diperlukan memory yang cukup besar juga, maka untuk mengolah data yang besar disarankan untuk menjalankan algoritma SPAM ini pada perangkat keras yang memiliki tingkat kinerja yang tinggi semisal komputer server.

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Pada zaman yang modern ini, perkembangan dunia bisnis sangatlah pesat. Salah satu bisnis yang kini mulai berkembang pesat yaitu bisnis ritel. Salah satu jenis perusahaan ritel yang menguasai pangsa pasar yaitu *hypermarket* dan perkulakan (grosir). Jenis perusahaan tersebut menjual barang-barang kebutuhan sehari-hari seperti bahan pangan. Oleh karena itu, tidak heran jika *hypermarket* atau grosir akan selalu dikunjungi oleh konsumen setiap harinya. Dengan maraknya berbagai macam perusahaan ritel yang ada, kini persaingan pun tak dapat dihindarkan. Tiap-tiap perusahaan tentunya mempunyai strategi yang berbeda-beda untuk menguasai pangsa pasar. Dalam pembuatan strategi pemasaran, tentu saja perusahaan harus dapat menganalisis dan mengevaluasi penjualan produk dari data transaksi yang ada. Salah satu faktor penting yang perlu diketahui oleh perusahaan yaitu adalah pola perilaku pembelian oleh konsumen dari perusahaan tersebut.

Perilaku pembelian konsumen tentunya akan sangat berpengaruh dalam kegiatan bisnis perusahaan. Seringkali perusahaan mengalami kesulitan dalam mempersiapkan barang dagangan karena tidak seimbang dengan tingkat permintaan (*demand*) dari konsumen. Misalnya yaitu saat ada perayaan hari raya tertentu, akan ada beberapa produk makanan yang tingkat penjualannya meningkat cukup tajam. Namun, terkadang perusahaan mengalami kendala karena kehabisan stok barang akibat kurang mengetahui produk apa saja yang sedang diminati konsumen pada saat itu. Oleh karena itu, penting bagi perusahaan untuk mengetahui pola pembelian konsumen sehingga perusahaan dapat mengetahui produk-produk apa saja yang perlu dipersiapkan dan diminati oleh konsumen.

Walaupun begitu, terdapat beberapa kendala untuk mengetahui pola dari para konsumen. Seiring berkembangnya waktu, pelanggan dari setiap perusahaan

bertambah setiap harinya. Kegiatan transaksi yang dilakukan pun meningkat setiap harinya sehingga mengakibatkan bertambah besarnya data transaksi yang ada pada perusahaan. Menganalisis data yang besar tidaklah mudah oleh karena beberapa penyebab. Salah satu contoh kendala yang dialami perusahaan yaitu lamanya waktu yang dibutuhkan untuk menganalisis. Apabila waktu yang digunakan terlalu lama, maka bisa jadi hasil dari analisis sudah tidak relevan dan tentunya tidak efisien terhadap waktu. Hal ini menjadi penghambat untuk menganalisis pola pembelian konsumen perusahaan tersebut.

Salah satu metode yang dapat digunakan untuk mengatasi masalah tersebut yaitu dengan menggunakan algoritma *Sequential Pattern Mining* (SPAM) untuk menganalisis pola pembelian konsumen dari data yang ada. Berdasarkan penelitian *Sequential Pattern Mining Using A Bitmap Representation* (Ayres, Gehrke, Yiu, & Flannick, 2002), algoritma SPAM terbukti lebih cepat dan efektif jika dibandingkan beberapa algoritma lainnya seperti SPADE dan *Prefix Span*. Dengan menggunakan algoritma ini, maka setiap perusahaan akan dapat mengetahui bagaimana pola pembelian dari konsumennya.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang ada, rumusan masalah yang akan dibahas dalam penelitian ini adalah bagaimana sistem dapat menganalisis pola pembelian konsumen perusahaan menggunakan algoritma *Sequential Pattern Mining*?

1.3 Batasan Masalah

Adapun batasan sistem dalam penelitian ini adalah sebagai berikut:

1. Data yang digunakan dalam pengujian algoritma SPAM pada penelitian ini merupakan data sederhana.
2. Data yang digunakan sebagai input berupa angka yang merupakan kode produk barang.

1.4 Tujuan Penelitian

Penelitian ini dilakukan dengan tujuan untuk menguji keakuratan dari algoritma *Sequential Pattern Mining* dalam membentuk pola pembelian konsumen pada perusahaan.

1.5 Metode Penelitian

Metodologi yang digunakan dalam penelitian ini adalah sebagai berikut:

1. Metode Studi Pustaka

Dalam melakukan penelitian ini, penulis melakukan studi pustaka dengan mempelajari teori-teori mengenai *Big Data*, *Data Mining*, *Data Warehouse* dan juga algoritma *sequential pattern mining SPAM*. Peneliti juga mempelajari teori yang digunakan dalam algoritma *SPAM* seperti *Lexicographic Tree for Sequences*, *Depth First Tree Traversal* dan *Pruning*.

2. Perancangan dan Pengembangan Sistem

Pada tahap ini, penulis melakukan perancangan sistem yang nantinya akan digunakan untuk menganalisis pola pembelian konsumen. Perancangan diawali dengan desain dan pembuatan sistem yang digunakan untuk pengolahan *database*, kemudian perancangan antarmuka. Nantinya, hasil *output* yang akan dikeluarkan oleh sistem yaitu berupa pola sekuensial konsumen perusahaan.

3. Evaluasi dan Analisis

Penulis melakukan evaluasi dan juga analisis terhadap pola sekuensial yang dihasilkan oleh sistem terhadap data uji. Pola sekuensial dari sistem dicocokkan dengan hasil pola yang dilakukan oleh penulis secara manual untuk dibandingkan.

1.6 Sistematika Penulisan

Sistematika laporan skripsi ini dibagi menjadi lima bab, sebagai berikut:

Bab I : Pendahuluan

Bab ini berisi mengenai latar belakang masalah, perumusan masalah, batasan masalah, tujuan penelitian, metodologi penelitian serta sistematika penulisan laporan skripsi.

Bab II : Tinjauan Pustaka

Pada bab ini berisi mengenai tinjauan pustaka dan landasan teori. Tinjauan pustaka menguraikan beberapa teori yang menjelaskan mengenai penggunaan algoritma *Sequential Pattern Mining* (SPAM) yang digunakan. Landasan teori berisi mengenai penjelasan tentang konsep dan juga prinsip utama yang diperlukan untuk memecahkan masalah pada penelitian.

Bab III : Analisis dan Perancangan Sistem

Bab ini berisi mengenai analisis dari teori-teori yang digunakan dalam penelitian dan mengimplementasikannya ke dalam sistem yang dibuat.

Bab IV : Implementasi dan Analisis Sistem

Berisi mengenai hasil implementasi dari sistem yang dibuat, beserta pembahasan dan juga analisis mengenai sistem tersebut.

Bab V : Kesimpulan dan Saran

Bab ini berisi mengenai kesimpulan dari hasil analisis dari implementasi sistem yang telah dibuat. Selain itu, bab ini juga berisi mengenai saran untuk pengembangan lebih lanjut mengenai sistem yang dibuat di masa mendatang.

BAB 5

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Dengan selesainya tugas akhir ini maka penulis dapat mengambil keputusan sebagai berikut:

1. Penggunaan algoritma SPAM dengan menggunakan *library* SPMF menghasilkan pola sekuensial yang valid
2. Pola sekuensial yang telah dibentuk dari sistem dapat digunakan untuk membantu perusahaan memprediksi pembelian konsumen di waktu yang akan datang.

5.2 Saran

Karena algoritma SPAM menelusuri setiap kemungkinan pada setiap *item* untuk mencari pola sekuensial, maka untuk pengolahan data yang besar dengan transaksi yang banyak sebaiknya menggunakan komputer yang memiliki kinerja yang tinggi.

DAFTAR PUSTAKA

- Ayres, J., Gehrke, J., Yiu, T., & Flannick, J. (2002). Sequential Pattern Mining using A Bitmap Representation.
- Liu, B. (2007). Web Data Mining. Heidelberg: Springer .
- Ho, J., Lukov, L., & Chawla, S. (2006). Sequential Pattern Mining with Constraints on Large Protein Database..
- Reshamwala, Alpa and Sunita M. Mahajan. (2014). An Early Pruning Algorithm for Sequential Pattern Mining.
- Mane, Rashmi V. and Dr. Vijay R.Ghorpade. (2013). Student's Performance Evaluation with SPAM Sequential Pattern Mining Algorithm.

©UKDWN