 http://dx.doi.org/10.29313/.v7i1.25474

Penerapan *Sequencial Pattern Mining* untuk Menemukan Pola Pembelian Konsumen Menggunakan Algoritma SPADE (*Sequencial Pattern Discovery using Equivalence Classes*)

Annajemin Rafiun*, Siti Sunendiari

Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Islam Bandung, Indonesia.

*annajeminr@gmail.com, diarisunen22@gmail.com

Abstract. During the last four years, according to Census data from the Central Statistics Agency (BPS), e-commerce in Indonesia has increased by 500 percent, but there are still many online businesses who don't know how to increase their company's profits. To solve this problem, data mining techniques are used. The purpose of this study is to find consumer purchasing patterns using the SPADE Algorithm (Sequencial Pattern Discovery using Equivalence Classes). The data used is secondary data which comes from the sales transaction data of the Jonas Sport Store, which amounts to 51 data and using data mining techniques and shows that from the search results of Frequent Sequence found 3 Frequent Sequences formed from goods E → F → M, namely consumers with sequence id 14 and 42 who are predicted to come to buy the same type of goods and also according to the order of the items on their next purchase. The item to be purchased is an item with the item code E → F → M, namely the Vapor Grip Gloves → Elbow pad → Jonas Eclipse.

Keywords: Algoritma SPADE, Sequential pattern mining, e-commerce, Frekuensi Sequence.

Abstrak. Selama kurun waktu empat tahun terakhir menurut data Sensus Badan Pusat Statistik (BPS), *e-commerce* di Indonesia mengalami peningkatan sebesar 500 persen, akan tetapi masih banyak pelaku bisnis *online* yang belum mengetahui cara untuk meningkatkan keuntungan perusahaanya. Untuk mengatasi masalah tersebut digunakan teknik data mining. Tujuan dari penelitian ini adalah untuk menemukan pola pembelian konsumen menggunakan Algoritma SPADE (*Sequencial Pattern Discovery using Equivalence Classes*). Data yang digunakan adalah data sekunder yang bersumber dari data transaksi penjualan Toko Jonas Sport yang berjumlah 51 data yang dianalisis menggunakan teknik data mining dan menunjukan bahwa dari hasil pencarian *Frequent Sequence* ditemukan *Frequent 3 Sequence* yang terbentuk dari barang E→ F→ M yaitu konsumen dengan *sequence id* 14 dan 42 yang diprediksi akan datang membeli jenis barang yang sama dan juga sesuai urutan item pada pembeliannya selanjutnya. Adapun barang yang akan dibeli adalah *item* dengan kode barang E → F →M yaitu Sarung tangan *Vapor Grip → Elbow pad → Jonas Eclipse.*

Kata Kunci: Algoritma SPADE, Sequential pattern mining, e-commerce, Frekuensi Sequence.

1. Pendahuluan

Selama kurun waktu empat tahun terakhir menurut Data sensus Badan Pusat Statistik (BPS), e-commerce di Indonesia mengalami peningkatan yang cukup besar yaitu sebanyak 500 persen. Mengutip data dari Global Web Index, Indonesia merupakan negara yang tingkat penggunan e-commerce tertinggi didunia yaitu pada tahun 2019. Hal ini disebabkan karena kebutuhan masyarakat yang terus menerus meningkat terhadap suatu produk, namun masih banyak pelaku bisnis yang kesulitan dalam mengatur strategi penjualan dengan cara yang tepat dan kesusahan mengolah data transaksi yang sangat banyak, perusahaan juga masih kesulitan untuk menentukan penyediaan stok barang dan keterkaitan antar barang yang dibeli oleh pembeli.

Permasalahan diatas dapat diatasi dengan menggunakan teknik data mining berbasis *sequential pattern mining.* Salah satu metode sequential pattern mining yaitu algoritma SPADE (Sequencial Pattern Discovery using Equivalence Classes) atau penemuan pola sekuensial menggunakan kelas yang ekivalen yang merupakan sebuah algoritma baru untuk menemukan pola sekuensial dalam data dengan cepat.

Toko JONAS adalah salah satu perusahaan dibidang e-commerce yang menjual perlengkapan futsal yang berlokasi di Kota Bandung, dimana setiap harinya transaksi penjualan dilakukan yang dapat menghasilkan kumpulan data transaksi penjualan yang cukup banyak, namun terkadang data tersebut hanya disimpan dan tidak diolah. Padahal jika data tersebut diolah dan dianalisis dengan baik dapat menghasilkan informasi mengenai data penjualan seperti pencatatan transaksi barang yang dibeli, barang terlaris, hubungan antara satu barang dengan barang yang lain. Informasi ini berguna untuk dijadikan bahan pengambilan keputusan oleh manajemen perusahaan. Oleh karena itu terkait dengan hal ini, penulis akan melihat Pola Sequencial Pattern Mining pembelian konsumen yang dilihat dari data transaksi penjualan di Toko JONAS SPORT sehingga dapat memberikan masukan baru untuk penjualan perusahaan dengan menggunakan teknik data mining. Berdasarkan uraian diatas, tujuan yang ingin dicapai dari penelitian ini adalah untuk mengetahui Pola pembelian barang dari data transaksi penjualan Toko JONAS SPORT dengan menggunakan Algorima Sequencial Pattern Minning using Equivalence Classes (SPADE).

2. Landasan Teori

Data *Mining*

Data *mining* adalah suatu teknik yang digunakan utuk menemukan informasi bermanfaat yang tersembunyi dalam suatu database yang sangat besar sehingga dapat ditemukan suatu pola yang menarik dari data yang belum diketahui sebelumnya. Contohnya dalam proses penemuan pola, hubungan atau tren baru dalam sebuah data dengan menggunakan teknik statistik dan matematika (Ardiansyah, 2013).

Pengelompokan Data Mining

Pada penelitian yang dilakukan oleh Ridwan (2013) ada beberapa kelompok dalam data mining berdasarkan tugas yang dilakukan, yaitu menemukan cara-cara untuk mendeskripsikan pola atau trend yang tersembunyi dalam data yang ada, estimasi yangi lebih ke arah numerik daripada kategori, prediksi hasilnya menunjukan sesuatu yang hasilnya belum terjadi atau mungkin terjadi di masa depan, klasifikasi yang variabel nya bersifat kategorik, pengklusteran adalah kumpulan record yang memiliki kemiripan satu dengan yang lain dan memiliki ketidakmiripan dengan record-record dalam cluster lain, dan asosiasi untuk mengidentifikasi hubungan antara berbagai peristiwa yang terjadi pada satu waktu.

Knowledge Discovery in Database (KDD)

Data mining merupakan bagian dari Knowledge Discovery in Database (KDD) yang berguna untuk mengestrak pola atau model dari data menggunakan algoritma yang spesisfik. Menurut Nofriyansyah (2014), istilah data mining dan Knowledge Discovery in Database (KDD) sering digunakan bergantian untuk mejelaskan proses penggalian informasi yang tersembunyi dalam suatu basis data yang besar . Salah satu proses tahapan dalam keseluruhan proses KDD adalah teknik data mining. Proses KDD adalah proses yang menggunakan database dalam jumlah yang besar, yang meliputi penyeleksian kumpulan data , preprocessing (cleaning), transformasi data, menerapkan metode data mining (algoritma) untuk menghitung pola data, lalu selanjutnya

mengevaluasi data yang terbentuk dari proses data mining sehingga ditemukan informasi yang baru dan bermanfaat (Fayyad, 1996).

Sequencial Pattern Mining

Penggalian pola sekuensial adalah pencarian semua sub*sequence* sub*sequence* berulang, yaitu sub*sequence* yang frekuensi kejadiannya lebih besar dari minimum support (Agrawal, 1995). Pola transaksi atau event yang terjadi dalam model proses bisnis biasanya tergambarkan dalam sebuah pola sekuensial. Pola sekuensial mengindikasikan bahwa transaksi biasanya terjadi secara serial terhadap waktu.

Algoritma SPADE

Langkah-langkah analisis dengan menggunakan algoritma SPADE dalam mencari *frequent sequence* kemudian menentukan rule dari *frequent sequence* tersebut adalah sebagai berikut (Zaki, 2001).

1. Menghitung *Frequent* 1 dilakukan pengecekan untuk setiap itemset dalam *sequence* database. Untuk masing- masing itemset, disimpan id-listnya (pasangan sid dan eid). Kemudian scan id-list dari masing-masing id-list tersebut, setiap ditemui sid yang sebelumnya ada atau sama maka nilai supportnya ditambah. *Sequence* yang dimasukkan dalam *frequent* 1-*sequence* adalah yang supportnya yang lebih dari min_sup.
2. Menghitung *Frequent* 2 *Sequence* yaitu data yang digunakan adalah data dari *frequent* 1-*sequence* sebelumnya, sehingga tidak perlu mencari dari *sequence* database lagi. Untuk setiap masing-masing *frequent* 1-*sequence* , gabungkan dengan semua *frequent* 1-*sequence* lainnya. Contohnya jika 1- *sequence* A digabungkan dengan 1-*sequence* B maka kemungkinan 2- *sequence* yang terjadi adalah A,B dimana A dan B muncul bersamaan dalam transaksi, A→ B dimana item B muncul setelah item A , dan B→ A dimana item B muncul setelah item A. Untuk setiap masing-masing penggabungan *frequent* 1-*sequence* ini dilakukan pengecekan apakah dalam id-listnya.
3. Menentukan *Frequent* K-*sequence*, untuk mencari *frequent* k-*sequence* ini dilakukan join pada *frequent* (k-1) *sequence* yang memiliki prefix yang sama. Contohnya untuk mencari 3-*sequence*, gabungkan *frequent sequence* dari 2-*sequence* yang memiliki prefix yang sama, untuk mencari 4-*sequence*, gabungkan *frequent sequence* dari 3-*sequence* yang memiliki prefix yang sama, dan seterusnya. Untuk mencari prefix *frequent* (k-1) *sequence*, hilangkan item terakhir dari *sequence* tersebut. Contoh, jika terdapat 4-*sequence* A→ $B \rightarrow C \rightarrow D$, maka prefix nya adalah A→ $B \rightarrow C$.

3. Hasil Penelitian dan Pembahasan

Data yang digunakan adalah data sekunder yang bersumber dari data hasil transaksi penjualan berbasis Online Shop Toko Jonas Sport yang berlokasi di Kota Bandung yang menjual peralatan Kiper untuk futsal. Periode data yang diambil yaitu hasil transaksi penjualan harian, yaitu pada bulan Januari Tahun 2020. Jumlah sampel yang digunakan dalam penelitian ini adalah sebanyak 51 sampel pembeli yang kemudian ditransformasikan menjadi format data *sequence* vertikal sehingga berjumlah 110 pasangan *sequence* id.

Data Deskriptif

Jenis barang yang paling banyak dibeli di Toko Jonas Sport adalah Fingertape yaitu sebanyak 13 item. Dan barang yang paling sedikit dibeli adalah Socktape yaitu hanya sebanyak 3 item.

Data Selection

Dalam penelitian ini, data yang digunakan merupakan data yang bersumber dari database transaksi penjualan di Toko Jonas Sport. Pada tahap ini akan dipilih atribut yang sesuai dengan kebutuhan algoritma Spade, yaitu atribut *Sid, Eid,* dan *Item* barang.

Data Preprocessing

1. Pembersihan Data (Data Cleaning)

Pembersihan data bertujuan untuk menghilangkan data transaksi penjualan dari kesalahan data dan membuang duplikasi data. Contohnya menghilangkan atribut nama yang

tidak digunakan karena sudah terwakili oleh SID (id pembeli),Event ID (urutan Kedatangan), Items (Barang) dan SIZE (banyak barang yang dibeli). Hal ini dikarenakan karena dalam Algoritma SPADE digunakan untuk mencari data-data yang memiliki urutan contohnya pada data transaksi.

2. Pengkategorian Data

Dalam penelitian ini tidak terjadi proses pengkategorian data , karena sesuai dengan tujuan yang akan dicapai setiap item (barang) tidak dapat disamakan atau digabungkan dalam satu kategori item lainnya. Proses pengkategorian data dilakukan jika data nya berjumlah besar dan bisa dikategorikan antar item. Pada penelitian ini item pertama dan item selanjutnya berbeda berbeda dan data yang digunakan dalam penelitian ini tidak terlalu besar dan dapat diatasi dengan penghalusan data saja pada data cleaning.

3. Data Transformtion

Transformasi data dilakukan dengan cara data di transformasi menjadi format data vertical. Database *sequence* menjadi berbentuk kumpulan urutan yang memiliki format [itemset: (*Sequence* ID,Event ID)].

Tabel 1. Format Data Vertikal SPADE

No	Sequence id	Even id	ID size	Item
1	1	1	1	A
2	1	2	1	B
3	2	1	1	R
⋮	⋮	⋮	⋮	⋮
10	6	1	1	J

Tabel diatas merupakan format data sequencial Pattern Minning dengan Algoritma SPADE. Setelah data telah sesuai dengan format Algoritma SPADE diperoleh, maka dapat di analisis untuk mencari Sequencial Patternnya.

4. *Sequencial Pattern Minning* dengan Algoritma SPADE

Setelah dibentuk tabel *sequence* vertikal , kemudian ditentukan minimum support yang bertujuan untuk mengetahui peluang kejadian beberapa item (barang) yang didebeli oleh pembeli dari keseluruhan total barang yang dibeli. Minimum support ditentukan oleh peneliti, dalam hal ini peneliti menggunakan Minimum support sebesar 50% atau 0.5. itemset dianggap *frequent* (sering muncul) jika muncul 50% dari keseluruhan id_pembeli yang digunakan yaitu muncul minimal 2 kali. Selanjutnya untuk menentukan 1-*Sequence*, dicari itemset yang muncul lebih banyak dari min_sup. Kemudian dibentuk tabel itemset tersebut lengkap dengan id_list,sid, dan eid. Itemset yang memiliki id_list lebih dari min_sup, dalam penelitian ini sebesar 50% dianggap memenuhi syarat support digunakan. Berikut merupakan tabel daftar frekuent 1-*Sequence* :

Tabel 2. Tabel Frequent 1-Sequence

A		B		E		F		H	
sid	eid	sid	Eid	sid	eid	sid	Eid	sid	Eid
1	1	1	2	14	1	14	2	14	3

18	1	18	2	42	1	42	2	42	3

Dari daftar 1-*Sequence* tersebut, kemudian digunakan untuk membentuk 2- *Sequence* . Caranya adalah dilakukan join untuk masing-masing *frequent* 1 *sequence* tersebut termasuk dengan dirinya sendiri. Contoh proses join : Apabila ada itemset yang dijoin dengan itemset A dijoin dengan itemset B, maka kemungkinan hasilnya *sequence* A, B yaitu itemset A dan B muncul bersamaan, sehingga *sequence* ini sama saja dengan B,A karena muncul bersamaan, sehingga hanya perlu dicek salah satunya saja. Hasil lain dari join *sequence* adalah *sequence* A → B dimana *sequence* ini berbeda dengan *sequence* B → A. *sequence* A → B menunjukkan itemset B muncul setelah itemset A, sedangkan *sequence* B → A menunjukkan itemset A muncul setelah itemset B. Karena dianggap *sequence* yang berbeda, maka keduanya harus dicek apakah id_Listnya memenuhi minimum support yang telah ditentukan.

Tabel 3. Tabel *Frequent 2 Sequence* A → B

A → B		
Sid	eid(A)	eid(B)
1	1	2
18	1	2

Dari Daftar *frequent* diatas, kita cek apakah id_list memenuhi minimum support atau tidak. *Sequence* A → B memenuhi syarat karena pada sid ke 1, itemset A muncul pada eid 1 dan itemset B muncul pada 2, atau dengan kata lain itemset B terjadi setelah itemset A. *Sequence* A → B juga terjadi pada sid 18 itemset A muncul pada eid 1 dan itemset B muncul pada eid 2. Jadi, *sequence* A → B dianggap *frequent* karena muncul 2 kali dari total 4 barang yang digunakan atau memiliki support 50%. Sedangkan B → A dianggap tidak *frequent* karena berdasarkan id_list A dan B, tidak ada eid A yang terjadi setelah B.

Tabel 4. Tabel *Frequent 2 Sequence* E → F

E → F		
Sid	eid(E)	eid(F)
14	1	2
42	1	2

Sequence E → F memenuhi syarat karena pada sid ke 14, itemset E muncul pada eid 1 dan itemset F muncul pada eid 2, atau dengan kata lain itemset F terjadi setelah itemset E. *Sequence* E → F juga terjadi pada sid 42 itemset E muncul pada eid 1 dan itemset B muncul pada eid 2. Jadi, *sequence* E → F dianggap *frequent* karena muncul 2 kali dari total 4 barang yang digunakan atau memiliki support 50%. Sedangkan F → E dianggap tidak *frequent* karena berdasarkan id_list E dan F, tidak ada eid E yang terjadi setelah F.

Tabel 5. Tabel *Frequent 2 Sequence* E → M

E → M		
Sid	eid(E)	eidM)
14	1	3
42	1	3

Sequence E → M memenuhi syarat karena pada sid ke 14, itemset E muncul pada eid 1 dan itemset M muncul pada eid 3, atau dengan kata lain itemset M terjadi setelah itemset E. *Sequence* E → M juga terjadi pada sid 42 itemset E muncul pada eid 1 dan itemset M muncul pada eid 2. Jadi, *sequence* E → M dianggap *frequent* karena muncul 2 kali dari total 4 barang yang digunakan atau memiliki support 50%. Sedangkan M → E dianggap tidak *frequent* karena berdasarkan id_list E dan F, tidak ada eid M yang terjadi setelah E.

Tabel 6. Tabel *Frequent 2 Sequence* F → M

F → M		
Sid	eid(F)	eid(M)
14	2	3
42	2	3

Sequence F → M memenuhi syarat karena pada sid ke 14, itemset E muncul pada eid 2 dan itemset M muncul pada eid 3, atau dengan kata lain itemset M terjadi setelah itemset F. *Sequence* F → M juga terjadi pada sid 42 itemset F muncul pada eid 2 dan itemset M muncul pada eid 3. Jadi, *sequence* F → M dianggap *frequent* karena muncul 2 kali dari total 4 barang yang digunakan atau memiliki support 50%. Sedangkan M → F dianggap tidak *frequent* karena berdasarkan id_list F dan M, tidak ada eid M yang terjadi setelah F.

Langkah berikutnya adalah menentukan *frequent* k-*sequence* menggunakan join pada semua (k-1) *sequence*. dari proses tersebut, didapatkan *sequence* dari perhitungan pada *sequence* E → F dengan *sequence* E → M. Berikut adalah *frequent* 3-*sequence* yang dapat dibentuk.

Tabel 7. Tabel *Frequent 3 Sequence* E → F → M

E→F → M			
Sid	eid(E)	eid(F)	eid(M)
14	1	2	3
42	1	2	3

Sequence E → F → M memenuhi syarat karena pada sid ke 14, itemset E muncul pada eid 1, itemset F muncul pada eid 2, dan itemset M muncul pada eid ke 3.itemset M terjadi setelah itemset F dan E.

Langkah berikutnya adalah mencari *frequent* 4 *sequence*, namun karena hanya ditemukan dengan 1 *frequent* 3 *sequence* saja, maka kemungkinan hanya *sequence* tersebut

digabungkan dengan *sequence* itu sendiri yang menghasilkan [E→ F →M], dimana *frequent sequence* ini dihentikan dan tidak dapat dibentuk.

Berdasarkan hasil id-list dari Tabel 3.7, Konsumen dengan *sequence* id 14 dan 42 diprediksi akan datang membeli jenis barang yang sama dan juga sesuai urutan item pada pembeliannya selanjutnya. Adapun barang yang akan dibeli adalah item dengan kode barang E → F →M yaitu Sarung tangan Vapor Grip → Elbow pad → Jonas Eclipse.

4. Kesimpulan

Berdasarkan pembahasan dalam penelitian ini, peneliti menyimpulkan beberapa hasil penelitian sebagai berikut:

Dalam perhitungan Algoritma Spade ditemukan 3 *Frequent Sequence*, *Frequent* 1 *Sequence* dibentuk oleh barang A (Jonas Helmet), B (Kneepad), E (Vapor Grip), F(Elbow Pad), dan H (Kaos Kaki Anti Slip) dengan pasangan urutan sid dan eidnya masing-masing.

Frequent 2 *Sequence*nya dijoin dari *frequent sequence* sebelumnya, *Frequent* 2 *Sequence* yang dibentuk adalah barang A → B, E → F, E→ M dan F → M dengan pasangan urutan sid dan eidnya masing-masing juga yang kemudian selanjutnya dijoin lagi untuk menemukan *Frequent* 3 *Sequence*nya.

Frequent 3 *Sequence* yang dapat dibentuk adalah dari barang E→ F→ M yaitu konsumen dengan *sequence* id 14 dan 42 yang diprediksi akan datang membeli jenis barang yang sama dan juga sesuai urutan item pada pembeliannya selanjutnya. Adapun barang yang akan dibeli adalah item dengan kode barang E → F →M yaitu Sarung tangan Vapor Grip → Elbow pad→ Jonas Eclipse.

5. Saran

Adapun saran dari hasil penelitian ini adalah sebagai berikut :

1. Untuk peneliti lain dapat menggunakan metode lain seperti prefixspan, GSP dan lainnya untuk membandingkan metode yang paling baik. dan juga diharapkan untuk menganalisis dengan menggunakan data yang cukup besar sehingga dapat mewakili dan menggambarkan pola pembelian yang lebih baik.
2. Untuk Toko Jonas Sport diharapkan dapat menggunakan data transaksi penjualan toko yang sangat banyak kemudian diolah menggunakan teknik data mining untuk menemukan informasi yang dapat digunakan untuk bahan pengambilan keputusan.

Daftar Pustaka

[1] Agrawal, R., & Srikant, R. 1995. Mining Sequencial Patterns. Proceedings the Eleventh International Conference on Data Engineering (pp. 3-14). doi; 10.1109/ICDE.1995.380415

[2] Ardiansyah, R. 2013. Sequencial Pattern Mining Pada Data Transaksi Penjualan Menggunakan Algoritma SPADE. Tugas Akhir Program Studi Ilmu Komputer Fakultas Teknologi Informatika dan Ilmu Komputer, Universitas Brawijaya Malang.

[3] Fayyad, U. 1996. Advances in Knowledge Discover and Data Mining. MIT Press.

[4] Juliasto, R., & Gunawan, D. 2015. *Sequential Pattern Mining Dengan Spade Untuk Prediksi Pembelian Spare Part Dan Aksesoris Komputer Pada Kedatangan Kembali Konsumen*. 314-325.

[5] Larose, D. T. 2005. Discovering Knowledge In Data: An Introduction to Data Mining. John Willey & Sons, Inc.Ardiansyah, R. (2013). Sequencial Pattern Mining Pada Data Transaksi Penjualan Menggunakan Algoritma SPADE. Tugas Akhir Program Studi Ilmu Komputer Fakultas Teknologi Informatika dan Ilmu Komputer, Universitas Brawijaya Malang.

[6] Nofriansyah, D. 2014. *Konsep Data Mining Vs Sistem Pendukung Keputusan*. Yogyakarta: Deepublish.

[7] Ridwan, M. 2013. Sistem Pendukung Keputusan Untuk Proses Kelulusan Dan Evaluasi Kinerja Akademik Mahasiswa Menggunakan Teknik Data Mining. Program Magister Teknik Elektro Fakultas Teknik Universitas Brawijaya Malang.

[8] Zaki, M. (2001). SPADE: An Efficient Algorithm For Mining *Frequent Sequence*s.

.