Pencarian *HighUtility Itemset* pada Dataset YooChoose Buys

R T Nugroho[*1], R Gunawan[2]

[1-2]Program Studi Informatika, Universitas Sanata Dharma, DIY, Indonesia

E-mail: ranggatrinugroho2430@gmail.com[1], rido@usd.ac.id[2]

Abstrak. Perkembangan *e-commerce* yang pesat, membuat strategi penjualan harus dioptimalkan untuk meningkatkan keuntungan bisnis. Dalam upaya untuk meningkatkan keuntungan, penting untuk mengidentifikasi pola pembelian pelanggan yang dapat memberikan keuntungan maksimal. *Data mining* menawarkan beberapa metode analisis pola pembelian, seperti *Frequent Itemset Mining* (FIM) dan *High utility Itemset* (HUI). FIM dapat menghasilkan pola yang kurang relevan bagi bisnis karena hanya mempertimbangkan kemunculan item daripada nilai ekonominya, metode HUI, khususnya algoritma HUI-Miner, digunakan untuk mengidentifikasi *itemset* yang mempunyai nilai keuntungan yang tinggi. HUI-Miner memanfaatkan struktur *utility-list* untuk meningkatkan efisiensi dalam pencarian pola pembelian yang menguntungkan. HUI-Miner diterapkan pada dataset transaksi YooChoose Buys untuk menganalisis pola pembelian pelanggan. Hasil analisis menunjukkan bahwa algoritma HUI-Miner mampu menemukan *itemset* yang bernilai tinggi secara efisien. Semakin tinggi *threshold* yang digunakan, semakin sedikit *itemset* yang memenuhi kriteria sebagai HUI, yang menunjukan bahwa pemilihan *threshold* yang tepat sangat berpengaruh terhadap hasil analisis. Hasil ini dapat dimanfaatkan untuk mengoptimalkan strategi pemasaran dan pengelolaan stok untuk meningkatkan keuntungan penjualan.

Kata kunci: E-commerce; *Data mining*; *High Utility Itemset*; HUI-Miner

Abstract. *The rapid development of e-commerce, makes sales strategies must be optimized to increase business profits. In an effort to increase profits, it is important to identify customer purchasing patterns that can provide maximum profit. In overcoming these problems, data mining offers several methods of analyzing purchasing patterns, such as Frequent Itemset Mining (FIM) and High utility Itemset (HUI). FIM can produce patterns that are less relevant to business because it only considers the occurrence of items rather than their economic value, HUI methods, especially the HUI-Miner algorithm, are used to identify itemsets that have high profit value. This algorithm can utilize the utility-list structure to improve efficiency in finding profitable purchase patterns. HUI-Miner is applied to the YooChoose Buys transaction dataset to analyze customer purchase patterns. The analysis results show that the HUI-Miner algorithm is able to find high-value itemsets efficiently. The higher the threshold used, the fewer itemsets that meet the criteria as HUI, which shows that the selection of the right threshold greatly affects the analysis results. These results can be used to optimize marketing strategies and stock management to increase sales profits.*

Keywords: *E-commerce; Data mining; High Utility Itemset; HUI-Miner*

1. Pendahuluan

E-Commerce, atau *electronic commerce*, adalah bentuk perdagangan elektronik yang memungkinkan penjualan barang dan jasa dilakukan secara *online* melalui media elektronik, khususnya internet. Perdagangan elektronik memainkan peran penting dalam memajukan teknologi informasi dan

komunikasi [1]. Penyebaran internet yang sangat besar telah mendorong pertumbuhan *e-commerce*, karena internet dan telepon sudah menjadi bagian dari kehidupan setiap orang [2]. Pertumbuhan ini dapat terlihat pada data dari YooChoose, yang menunjukkan aktivitas pembelian pengguna dalam bisnis *e-commerce* besar di Eropa selama periode enam bulan, khususnya pada platform yang menjual berbagai kebutuhan konsumen [3].

Seiring dengan pertumbuhan yang pesat ini, penting untuk mengoptimalkan penjualan agar bisnis tetap kompetitif. Dalam upaya tersebut terdapat beberapa metode analisis pola pembelian yang umum digunakan dalam *transactional database*, seperti *Frequent Itemset Mining* (FIM) dan *High utility itemset* (HUI) yang sudah banyak dipakai untuk permasalahan yang berkaitan dengan jumlah pembelian barang dan keuntungan dari barang yang dibeli [4].

Metode *Frequent Itemset Mining* (FIM) memiliki kelebihan dalam mengurangi ruang pencarian, sehingga dapat mengurangi jumlah kombinasi yang perlu dieksplorasi namun, kelemahan dari metode ini adalah bahwa FIM hanya mempertimbangkan frekuensi kemunculan *itemset* tanpa mempertimbangkan nilai atau *profitabilitas* setiap item. Akibatnya, metode ini dapat menghasilkan *itemset* yang tidak relevan atau tidak menguntungkan [4]. Dalam kehidupan nyata, pola yang sering muncul tidak selalu merupakan pola yang paling menarik atau berguna [5].

Metode *High utility itemset* (HUI) dapat mengatasi kelemahan yang terdapat pada metode *Frequent Itemset Mining* (FIM) karena HUI memungkinkan analisis yang lebih relevan dalam konteks bisnis dengan mempertimbangkan nilai atau keuntungan dari setiap item, bukan hanya frekuensinya saja [4]. HUI-Miner, dirancang untuk meningkatkan efisiensi dalam menemukan *itemset* dengan utilitas tinggi. Algoritma ini menggunakan struktur *utility-list* untuk menyimpan informasi utilitas dan *heuristik*, HUI-Miner dapat menghitung utilitas *itemset* dengan lebih cepat tanpa perlu memindai ulang database. Hal ini sangat menguntungkan ketika bekerja dengan dataset besar, di mana pemindaian ulang dapat menjadi sangat mahal [6].

Penelitian ini bertujuan untuk melihat kinerja dari algoritma HUI-Miner dengan menggunakan data transaksi dari *e-commerce* YooChoose untuk mengidentifikasi pola pembelian yang bernilai tinggi. Dengan menerapkan algoritma ini diharapkan dapat ditemukan pola transaksi yang tidak hanya sering terjadi tetapi juga memberi keuntungan optimal bagi bisnis.

2. Kerangka Teoritis

2.1. E-Commerce

E-commerce adalah jenis transaksi bisnis seperti distribusi, pembelian, penjualan, dan pelayanan yang dilakukan secara elektronik melalui jaringan komputer, terutama internet, tetapi juga jaringan eksternal. Akses internet, kemudahan mendapatkan informasi, kemampuan tenaga kerja, dan tanggung jawab manajemen informasi adalah beberapa dimensi atau indikator *e-commerce*. Dimensi ini mencakup proses, pemasaran, dan pembayaran [7]. Istilah *e-commerce* secara umum digunakan untuk menggambarkan penjualan barang dan jasa melalui internet [8].

Pada tahun 2015 RecSys mengadakan kompetisi atau *challenge* yang berfokus pada isu-isu yang berkaitan dengan sistem rekomendasi. Kompetisi tersebut menggunakan data yang disediakan oleh YooChoose, sebuah perusahaan yang menyediakan layanan di bidang *e-commerce*. Dataset yang disediakan oleh YooChoose terbagi menjadi dua bagian, bagian pertama berisi tentang *training data file*, bagian kedua berisi tentang *test data file* [9].

2.2. Data mining

Data mining adalah proses mengekstraksi dan mengidentifikasi informasi terkait dari berbagai *database* besar menggunakan metode statik, matematika, kecerdasan buatan (AI), dan pembelajaran mesin [10]. *Data mining* secara khusus adalah suatu proses ekstraksi atau penggalian data dan informasi yang besar, yang belum diketahui sebelumnya, namun dapat dipahami dan berguna dari *database* yang besar serta digunakan untuk membuat suatu keputusan bisnis yang sangat penting [11]. *Data mining* telah berkembang menjadi alat bantu untuk menemukan pola penting dalam basisdata yang sangat besar, yang tidak dapat ditemukan secara manual [10].

2.3. High Utility Itemset (HUI)

2.3.1. Pengertian HUI. *High Utility Itemset* adalah *itemset* yang memenuhi ambang batas minimum utilitas yang ditentukan oleh pengguna. Artinya, HUI adalah kombinasi item yang tidak hanya sering muncul dalam transaksi, tetapi juga memberikan nilai yang signifikan ketika dipertimbangkan bersama-sama [12]. HUI merupakan masalah penambangan data yang populer karena mempertimbangkan faktor utilitas, seperti kuantitas dan keuntungan unit item selain ukuran frekuensi dari basis data *transactional*, hal ini membantu menemukan produk/item yang sulit dilacak hanya dengan menggunakan *frequent itemset [13]*.

2.3.2. Tujuan High Utility Itemset. Tujuan utama dari HUI yaitu, untuk mengidentifikasi kombinasi item yang memberikan nilai utilitas tinggi dalam transaksi. Hal ini membantu dalam memahami *itemset* mana yang paling menguntungkan bagi bisnis atau organisasi. Dengan mengetahui *itemset* yang memiliki utilitas tinggi, perusahaan dapat membuat keputusan yang lebih baik terkait strategi pemasaran, pengelolaan inventaris, dan penawaran produk. Ini memungkinkan perusahaan untuk fokus pada item yang memberikan kontribusi terbesar terhadap pendapatan. Dengan fokus pada *itemset* yang memberikan utilitas tinggi, perusahaan dapat meningkatkan profitabilitas mereka dengan mengoptimalkan penawaran produk dan mengurangi biaya yang terkait dengan item yang kurang menguntungkan [12].

2.3.3. Langkah-langkah Menemukan High Utility Itemset (HUI).

Langkah-langkah berikut menjelaskan proses untuk menemukan HUI dengan menggunakan algoritma HUI-Miner:

1. Persiapan data, mengumpulkan dan menyiapkan data transaksi yang berisi item-item yang dibeli dalam setiap transaksi.
2. Hitung utilitas dasar, hitung utilitas dasar setiap item dalam basis data dengan mengalikan harga per unit dan jumlah item yang terjual. Perhitungan utilitas diperoleh dengan menggunakan persamaan:

$$u(i_j, T_q) = q(i_j, T_q) \times pft(i_j) \quad (1)$$

   Keterangan:
   q(ij,Tq) = kuantitas dari item ij dalam transaksi Tq. Kuantitas menunjukkan jumlah item ij yang terdapat dalam transaksi tersebut.
   pft(ij) = profit per unit dari item ij. Profit per unit adalah nilai atau harga yang diperoleh dari setiap unit item.
3. Tentukan ambang batas, tentukan batas minimum *utility* (minutil). Hanya *itemset* yang memiliki *utility* lebih dari batas ini yang akan dianggap sebagai HUI.
4. Membuat *utility list*, untuk setiap item, buat daftar utilitas. Daftar utilitas membantu dalam proses *pruning* karena menyimpan informasi tentang utilitas dari *itemset* yang relevan.
5. *Pruning* Awal, mengurangi jumlah item yang perlu dievaluasi, lakukan *pruning* awal untuk menyingkirkan item yang utilitasnya lebih rendah dari ambang batas.

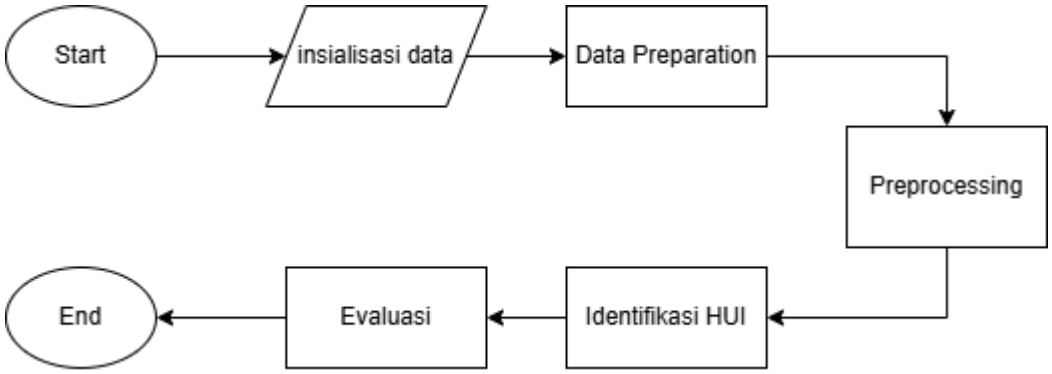
$$HUI \leftarrow \{X | u(X) \geq TU \times \delta\} \quad (2)$$

   Keterangan:
   u(X) = utilitas dari *itemset* X. Utilitas ini dihitung berdasarkan nilai dan kuantitas item dalam *itemset* tersebut.
   TU = total utilitas dari seluruh basis data
   δ = ambang batas minimum (minutil) yang ditetapkan untuk menentukan apakah *itemset* tersebut dianggap sebagai HUI

6. *Generate Candidate Itemsets*, HUI-Miner membuat kandidat item dari *itemset* yang sudah ada. Proses ini menggabungkan *itemset* yang sudah ada melalui operasi penggabungan (*join operation*) untuk membentuk *itemset* baru.
7. Hitung *utility* untuk kandidat, hitung *utility* untuk setiap kandidat *itemset* yang dihasilkan. Bandingkan *utility* ini dengan *threshold* (ambang batas) yang telah ditentukan.
8. Identifikasi HUI, *itemset* yang utilitasnya lebih besar dari ambang batas disebut HUI dan harus disimpan untuk analisis lebih lanjut.
9. Iterasi, ulangi proses ini untuk *itemset* yang lebih besar hingga tidak ada kandidat baru yang dapat dihasilkan.
10. Evaluasi hasil, setelah menemukan HUI, evaluasi hasilnya untuk memastikan bahwa *itemset* yang ditemukan relevan dan bermanfaat.
11. *Output* HUI, Hasilnya adalah daftar HUI yang ditemukan, yang dapat digunakan untuk analisis dan pengambilan keputusan bisnis berikutnya.
    [13]

3. Metodologi Penelitian

Metodologi yang digunakan dalam penelitian ini terbagi menjadi beberapa tahap seperti yang ditunjukkan pada Gambar 1.



Gambar 1. Tahapan Penelitian

Tahapan penelitian ini dibagi menjadi beberapa langkah yaitu:

1. Inisialisasi Data
   Insialisasi data adalah tahap pertama penelitian. Pada tahap ini, peneliti menentukan subjek penelitian, mengumpulkan data dari sumber yang relevan, dan memilih atribut awal yang paling relevan untuk dianalisis.
2. Persiapan Data (*Data Preparation)*
   Tahap selanjutnya adalah *data preparation*, yaitu mempersiapkan data dari hasil inisialisasi data agar bisa diproses dan di analisa lebih lanjut. Pada tahap ini, data diperiksa untuk memastikan konsistensi, dibersihkan dari duplikat, menghitung *utility* per item, dan disusun dalam format yang tepat agar dapat diolah pada tahap berikutnya.
3. *Preprocessing*
   Tahap selanjutnya adalah *preprocessing*, di mana data dibuat lebih siap untuk dianalisis lebih lanjut. Tujuan dari *preprocessing* adalah untuk membersihkan, menormalkan, dan mempersiapkan data sehingga dapat diolah lebih efektif oleh algoritma analisis atau model *machine learning* [14]. Pada tahap ini, beberapa proses dilakukan, seperti menghapus nilai yang hilang (NaN) atau *null*, melakukan *feature selection* untuk meringankan beban pemrosesan dalam model *data mining* [15]. Selain itu, *feature selection* digunakan untuk memilih fitur yang relevan dan menghilangkan fitur yang tidak relevan, dan bagian terakhir yaitu melakukan

grouping atau mengelompokkan data berdasarkan atribut seperti *Session ID* untuk memudahkan analisis lebih lanjut.

4. Identifikasi *High Utility Itemset* (HUI)
   Setelah data siap, dilakukan analisis untuk mengidentifikasi *high utility itemset* (HUI) menggunakan algoritma *HUI-Miner*. Tujuannya adalah untuk menemukan kombinasi item yang dapat memberikan keuntungan tinggi.
5. Evaluasi
   Pada tahap evaluasi, hasil *high utility itemset* dianalisis serta disaring berdasarkan *threshold* (batas nilai) yang telah ditetapkan. *Itemset* yang dihasilkan akan di evaluasi berdasarkan nilai dan manfaatnya dalam mendukung pengambilan keputusan bisnis, terutama dalam meningkatkan keuntungan penjualan.

Penelitian ini bertujuan untuk memperoleh *High Utility Itemset* (HUI) dengan menggunakan model HUI-Miner. Model ini diterapkan untuk menemukan *itemset* dengan nilai utilitas tinggi berdasarkan data transaksi dari YooChoose Buys. Model ini digunakan karena mempunyai keuntungan dalam menyimpan data transaksi dalam bentuk *utility-list*, sehingga dapat meningkatkan efisiensi dalam proses pencarian HUI.

4. Hasil dan Pembahasan

4.1. Data Penelitian

Data yang digunakan dalam penelitian ini merupakan data transaksi dari platform *e-commerce* YooChoose. Data ini mencakup lebih dari 1 juta data transaksi dengan lima atribut yang dapat dianalisis lebih lanjut. Namun, data mentah tersebut masih terdapat banyak data yang tidak penting dan atribut yang kurang mendukung dalam proses analisis. Oleh karena itu, data perlu disiapkan dengan baik, dan diperlukan proses pengolahan data awal untuk mendapatkan data yang berkualitas sebelum digunakan dalam penelitian. Berikut ini merupakan data mentah dari data transaksi *e-commerce* YooChoose Buys.

Tabel 1. Tabel *Dataset* Transaksi

Session ID	Time Stamp	Item ID	Item Price	Quantity
420374	2014-04-06T18:44:58.314Z	214537888	12462	1
420374	2014-04-06T18:44:58.325Z	214537850	10471	1
281626	2014-04-06T09:40:13.032Z	214535653	1883	1
420368	2014-04-04T06:13:28.848Z	214530572	6073	1
420368	2014-04-04T06:13:28.848Z	214530572	6073	1
140806	2014-04-07T09:22:28.132Z	214668193	523	1
140806	2014-04-07T09:22:28.176Z	214587399	1046	1
140806	2014-04-07T09:22:28.219Z	214586690	837	1
140806	2014-04-07T09:22:28.268Z	214774667	1151	1
140806	2014-04-07T09:22:28.280Z	214578823	1046	1

Karena data memiliki jumlah 1 juta data dengan 5 atribut yang terlalu banyak maka dalam pengujian hanya akan menggunakan 1.000 *sample* data dan beberapa *sample* ter tampil seperti pada Tabel 1. Berikut ini merupakan penjelasan dari atribut yang akan digunakan:

1. *Session ID* merupakan kode sesi pengguna.
2. *Time Stamp* merupakan waktu dibelinya suatu item.
3. *Item ID* merupakan *id* dari suatu produk atau item.
4. *Item Price* merupakan harga dari suatu produk.
5. *Quantity* merupakan jumlah dari barang yang di beli.

4.2. Persiapan Data (Data Preparation)

Data yang telah dibaca harus disiapkan sebelum diproses karena terdapat duplikasi, data tidak penting, dan data kosong yang dapat membuat data tidak valid atau bahkan tidak dapat diproses.

4.2.1 Pembersihan Data

Tabel 2 menunjukkan hasil dari data mentah yang telah mengalami proses pembersihan data.

Tabel 2. Tabel Hasil Pembersihan Data

Session ID	Time Stamp	Item ID	Item Price	Quantity
420374	2014-04-06T18:44:58.314Z	214537888	12462	1
420374	2014-04-06T18:44:58.325Z	214537850	10471	1
281626	2014-04-06T09:40:13.032Z	214535653	1883	1
420368	2014-04-04T06:13:28.848Z	214530572	6073	1
420368	2014-04-04T06:13:28.848Z	214835025	2617	1
140806	2014-04-07T09:22:28.132Z	214668193	523	1
140806	2014-04-07T09:22:28.176Z	214587399	1046	1
140806	2014-04-07T09:22:28.219Z	214586690	837	1
140806	2014-04-07T09:22:28.268Z	214774667	1151	1
140806	2014-04-07T09:22:28.280Z	214578823	1046	1

Hasil yang di dapatkan dari Tabel 2 menunjukkan bahwa dalam proses pembersihan data, terutama penghapusan data duplikat, terlihat pada Tabel 1 bahwa *Session ID* (420368) mempunyai *Time Stamp*, *Item ID*, dan *Item Price* yang sama. Oleh karena itu, salah satu data duplikat tersebut akan dihapus.

4.2.2. Menghitung Utility Per Item

Utility setiap item didapatkan dari mengalikan *item price* dengan *quantity*. Tabel 3 menunjukkan *utility* masing-masing item yang ada.

Tabel 3. Tabel *Utility* per *item*

Session ID	Time Stamp	Item ID	Item Price	Quantity	*Utility*
420374	2014-04-06T18:44:58.314Z	214537888	12462	1	12462
420374	2014-04-06T18:44:58.325Z	214537850	10471	1	10471
281626	2014-04-06T09:40:13.032Z	214535653	1883	1	1883
420368	2014-04-04T06:13:28.848Z	214530572	6073	1	6073
420368	2014-04-04T06:13:28.848Z	214835025	2617	1	2617
140806	2014-04-07T09:22:28.132Z	214668193	523	1	523
140806	2014-04-07T09:22:28.176Z	214587399	1046	1	1046
140806	2014-04-07T09:22:28.219Z	214586690	837	1	837
140806	2014-04-07T09:22:28.268Z	214774667	1151	1	1151
140806	2014-04-07T09:22:28.280Z	214578823	1046	1	1046

4.3. Preprocessing

4.3.1. Menghapus nilai null

Data yang mengandung nilai yang kosong atau null akan dihapus secara langsung, sehingga hanya tersisa data yang lengkap dan sudah siap untuk di proses lebih lanjut.

4.3.2. Feature Selection

Tabel 4 menunjukkan hasil setelah atribut yang tidak relevan, seperti *Time Stamp* dihapus.

Tabel 4. Tabel Hasil *Feature Selection*

Session ID	Item ID	Item Price	Quantity	*Utility*
420374	214537888	12462	1	12462
420374	214537850	10471	1	10471
281626	214535653	1883	1	1883
420368	214530572	6073	1	6073
420368	214835025	2617	1	2617
140806	214668193	523	1	523
140806	214587399	1046	1	1046
140806	214586690	837	1	837
140806	214774667	1151	1	1151
140806	214578823	1046	1	1046

Kolom *Time Stamp* dihapus karena tidak memiliki relevansi dalam menemukan HUI.

4.3.3. Grouping

Data transaksi yang sudah selesai di *preprocessing*, kemudian akan di kelompokan berdasarkan *Session ID* yang sama.

Tabel 5. Tabel Hasil *Grouping*

Session ID	Item ID	Item Price	*Quantity*	*Utility*
420374	214537888, 214537850	12462, 10471	1, 1	12462, 10471
281626	214535653	1883	1	1883
420368	214530572, 214835025	6073, 2617	1, 1	6073, 2671
140806	214668193, 214587399, 214586690, 214774667, 214578823	523, 1046, 837, 1151, 1046	1, 1, 1, 1, 1	523, 1046, 837, 1151, 1046

4.4. Identifikasi HUI

Data yang sudah siap dan valid akan digunakan untuk mengidentifikasi *High Utility Itemset* (HUI). Proses ini dilakukan dengan menerapkan algoritma HUI-Miner, untuk mencari *itemset* dengan kualitas yang tinggi. Setiap *itemset* yang memenuhi *threshold* (ambang batas nilai) yang sudah di tentukan akan dianggap sebagai HUI. Hasil HUI ditampilkan pada Tabel 6. Dari 1.000 data yang diambil dan sudah diolah, didapatkan total *utility* sebesar 4.185.798. Dari total *utility* tersebut, diambil *threshold* (minutil) sebesar 1%, yaitu 41.857,98, sebagai nilai minutil untuk menentukan *itemset* yang memenuhi kriteria sebagai *High Utility Itemset* (HUI).

Tabel 6. Hasil Akhir HUI

Itemset	*Utility*
214587765	276985
214840762	116868
214821277	106692
214587898	104510
214821285, 214826803	81640
214696434	62622
214567404, 214821277, 214821371, 214835585	61147
214567404, 214821277, 214835585	60101
214821277, 214821371, 214835585	59577
214839313	58632
214826803	58632
214821290	58576
214821277, 214835585	58531
214835585	52255
214539845	52150
214600502	52150
214821285	48116
214716932	46486
214833802	45972
214594678	44500
Total *Utility*	1506142

Hasil dari Tabel 6 menunjukkan bahwa semakin tinggi *threshold* (ambang batas nilai) yang digunakan, semakin sedikit jumlah *itemset* yang ditemukan sebagai *High Utility Itemset* (HUI). Dari Tabel 6, total *utility* yang diperoleh sebesar 1.506.142, dengan *itemset* yang memiliki *utility* tertinggi yaitu 214587765 sebesar 276.985.

Itemset yang mempunyai *utility* yang tinggi menunjukkan bahwa produk atau kombinasi produk tersebut memiliki kontribusi yang besar terhadap total keuntungan. Sebagai contoh, item 214821285 dan 214826803 membentuk kombinasi dengan *utility* yang diperoleh sebesar 81640, hal ini menunjukkan adanya hubungan yang kuat antara kedua item dalam meningkatkan keuntungan penjualan.

Jumlah *itemset* yang teridentifikasi sebagai HUI juga dipengaruhi oleh *threshold* yang digunakan. Semakin rendah *threshold*, semakin banyak *itemset* yang memenuhi sebagai kriteria HUI, sedangkan semakin tinggi *threshold*, semakin sedikit *itemset* yang teridentifikasi sebagai HUI, karena hanya *itemset* dengan nilai *utility* yang sangat tinggi yang akan dipertahankan dalam analisis. Hal ini menunjukkan bahwa pemilihan *threshold* yang tepat sangat penting dalam menemukan pola pembelian yang paling menguntungkan.

5. Kesimpulan

Berdasarkan hasil identifikasi HUI, dapat disimpulkan bahwa:

1. Algoritma HUI-Miner mampu menemukan *itemset* yang memenuhi *threshold* untuk menjadi *High Utility Itemset* (HUI), dengan efisiensi tinggi. Hal ini memungkinkan identifikasi pola pembelian yang dapat memberikan keuntungan secara signifikan.
2. Pemilihan *threshold* sangat berpengaruh terhadap jumlah *itemset* yang teridentifikasi sebagai HUI. Semakin rendah *threshold*, semakin banyak *itemset* yang memenuhi sebagai kriteria HUI, sedangkan semakin tinggi *threshold*, semakin sedikit *itemset* yang teridentifikasi sebagai HUI.
3. Hasil analisis *High Utility Itemset* (HUI) dapat digunakan untuk menentukan produk atau kombinasi produk yang memiliki dampak ekonomi terbesar bagi bisnis. Misalnya, bisnis dapat mengoptimalkan strategi penjualan mereka dengan memahami *itemset* dengan nilai *utility* tinggi untuk meningkatkan pemasaran, memberikan saran untuk strategi pemasaran berbasis data, atau menggunakannya untuk pengelolaan stok yang lebih efektif untuk memastikan bahwa produk memiliki nilai *utility* tinggi.
4. *Itemset* dengan *utility* yang tinggi menunjukkan produk atau kombinasi produk yang berkontribusi besar terhadap total keuntungan. Misalnya, item 214587765 memiliki *utility* tertinggi sebesar 276.985, yang menunjukkan bahwa produk ini memiliki nilai ekonomi yang signifikan dalam transaksi. Selain itu, kombinasi produk seperti 214821285 dan 214826803 dengan *utility* 81.640 menunjukkan hubungan yang kuat dalam meningkatkan keuntungan penjualan.

6. Referensi

[1] G. Taher, "E-commerce: advantages and limitations," *International Journal of Academic Research in Accounting Finance and Management Sciences*, vol. 11, no. 1, pp. 153–165, 2021.

[2] V. Jain, B. Malviya, and S. Arya, "An overview of electronic commerce (e-Commerce)," *The journal of contemporary issues in business and government*, vol. 27, no. 3, pp. 665–670, 2021.

[3] O. Mokryn, V. Bogina, and T. Kuflik, "Will this session end with a purchase? Inferring current purchase intent of anonymous visitors," *Electron Commer Res Appl*, vol. 34, p. 100836, 2019, doi: https://doi.org/10.1016/j.elerap.2019.100836.

[4] Y. Djenouri, P. Fournier-Viger, A. Belhadi, and J. Chun-Wei Lin, "Metaheuristics for frequent and high-*utility itemset* mining," *High-Utility Pattern Mining: Theory, Algorithms and Applications*, pp. 261–278, 2019.

[5] Q.-H. Duong, P. Fournier-Viger, H. Ramampiaro, K. Nørvåg, and T.-L. Dam, "Efficient high *utility itemset* mining using buffered *utility*-lists," *Applied Intelligence*, vol. 48, pp. 1859–1877, 2018.

[6] J.-F. Qu, M. Liu, and P. Fournier-Viger, "Efficient Algorithms for High *Utility Itemset* Mining Without Candidate Generation," in *High-Utility Pattern Mining: Theory, Algorithms and Applications*, P. Fournier-Viger, J. C.-W. Lin, R. Nkambou, B. Vo, and V. S. Tseng, Eds., Cham: Springer International Publishing, 2019, pp. 131–160. doi: 10.1007/978-3-030-04921-8_5.

[7] R. Teresya, R. R. Nabiilah, and S. Tunnajah, "Literature Review E-Commerce: Profitabilitas, tekanan eksternal dan kemudahan pengguna," *Jurnal Ekonomi Manajemen Sistem Informasi*, vol. 3, no. 4, pp. 474–484, 2022.

[8] Y. L. R. Rehatalanit, "Peran e-commerce dalam pengembangan bisnis," *Jurnal Teknologi Industri*, vol. 5, 2021.

[9] J. A. Napieralski, "Statistical methods for data prediction," in *2016 XII International Conference on Perspective Technologies and Methods in MEMS Design (MEMSTECH)*, 2016, pp. 120–123. doi: 10.1109/MEMSTECH.2016.7507532.

[10] A. Yani, Z. Azmi, and D. Suherdi, "Implementasi *Data mining* Menganalisa Data Penjualan Menggunakan Algoritma K-Means Clustering," *Jurnal Sistem Informasi Triguna Dharma (JURSI TGD)*, vol. 2, no. 2, pp. 315–323, 2023.

[11] C. Zai, "Implementasi *data mining* sebagai pengolahan data," *Jurnal Portal Data*, vol. 2, no. 3, 2022.

[12] B. Rahmati and M. K. Sohrabi, "A Systematic Survey on High *Utility Itemset* Mining," *Int. J. Inf. Technol. Decis. Mak.*, vol. 18, pp. 1113–1185, 2019, [Online]. Available: https://api.semanticscholar.org/CorpusID:145922793

[13] J. M.-T. Wu, J. C.-W. Lin, and A. Tamrakar, "High-*Utility Itemset* Mining with Effective Pruning Strategies," *ACM Trans. Knowl. Discov. Data*, vol. 13, no. 6, Nov. 2019, doi: 10.1145/3363571.

[14] F. Putra, H. F. Tahiyat, R. M. Ihsan, R. Rahmaddeni, and L. Efrizoni, "Penerapan Algoritma K-Nearest Neighbor Menggunakan Wrapper Sebagai Preprocessing untuk Penentuan Keterangan Berat Badan Manusia: Application of K-Nearest Neighbor Algorithm Using Wrapper as Preprocessing for Determination of Human Weight Information," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 1, pp. 273–281, 2024.

[15] R. Sowmya and K. R. Suneetha, "*Data mining* with big data," in *2017 11th International Conference on Intelligent Systems and Control (ISCO)*, IEEE, 2017, pp. 246–250.