Implementasi Data Mining Untuk Product Bundling Pada Coffee Shop

1st Abdurrahman Aziz
Fakultas Rekayasa Industri
Universitas Telkom
Bandung, Indonesia
abduraziz@student.telkomuniversity.ac.id

2nd Faqih Hamami
Fakultas Rekayasa Industri
Universitas Telkom
Bandung, Indonesia
faqihhamami @telkomuniversity.ac.id

3rd Iqbal Yulizar
Fakultas Rekayasa Industri
Universitas Telkom
Bandung, Indonesia
iqbalyulizar@telkomuniversity.ac.id

***Abstrak*— Di era sekarang dalam menjalankan bisnis coffee shop sudah terbantu dengan layanan pengantaran makanan dan minuman, seperti Gojek dan Grab. Dalam menjalankan bisnis coffee shop perlu merancang strategi agar penjualan produk meningkat. Strategi pemasaran yang umum dilakukan yaitu menggunakan product bundle. Product bundle dilakukan dengan menjual dua produk atau lebih dalam 1 paket, biasanya disertai dengan potongan harga. Dalam penelitian ini, penulis melakukan data mining pada data penjualan milik Authen Café & Space untuk menghasilkan product bundle. Penelitian ini menggabungkan clustering dan association rules mining dalam menghasilkan product bundle. Hasil dari proses clustering diperoleh cluster efektif sebanyak tiga cluster. Setiap cluster kemudian dilakukan association rules mining. Association rules mining menggunakan algoritma fp-growth dengan lift ratio minimal bernilai 1 pada dataset cluster pertama menghasilkan 6 rules, dataset cluster kedua menghasilkan 4 rules dan dataset cluster terakhir menghasilkan 8 rules. Berdasarkan hasil association rules mining maka saran product bundle yang dapat diterapkan dengan mengambil nilai confidence dan lift tertinggi pada hasil association rules mining setiap cluster.**

***Kata kunci* — Product bundle, Clustering, Association rules, K-means, Fp-growth**

I. PENDAHULUAN

Bisnis coffee shop di Indonesia saat ini menjadi bisnis yang naik daun. Menurut penelitian yang dilakukan Toffin dengan majalah Mix menunjukkan hingga tahun 2019 terjadi pertumbuhan gerai *coffee shop* di Indonesia hinga 3 kali lipat selama 3 tahun terakhir. Peningkatan gerai *coffee shop* tersebut sejalan dengan peningkatan jumlah konsumsi domestik kopi di Indonesia setiap tahunnya berdasarkan data dari International Coffee Organization. Pertumbuhan gerai *coffee shop* yang terjadi menyebabkan semakin ketatnya persaingan bisnis.

Di era sekarang, penjualan di *coffee shop* sudah terbantu dengan penjualan *online* melalui layanan *ride-hailing*. Hasil survei yang dilakukan Rakuten Insight pada tahun 2023 dari 9874 responden di Indonesia menunjukkan bahwa 75% responden menggunakan layanan Gofood dan 57% responden menggunakan Grabfood.

Strategi pemasaran yang sering digunakan di layanan Gofood dan Grabfood yaitu menggunakan *product bundle*. Menurut laporan dari gojek di tahun 2020 menunjukkan bahwa pesanan *bundle* memberikan kontribusi sebesar 40% dari semua transaksi Gofood.

Di Tengah ketatnya persaingan usaha *coffee shop,* permasalahan umum yang dihadapi oleh pemilik *coffee shop* yaitu bagaimana cara meningkatkan penjualan. Penulis menawarkan Solusi dengan mengoptimalkan penjualan secara *online*, disamping penjualan secara *dine-in* dengan memberikan promo *bundle.*

Untuk membuat promo *bundle*, peneliti perlu mengetahui perilaku pelanggan yang melakukan transaksi di *coffee shop*. Untuk mengelompokkan pelanggan berdasarkan pola perilaku ketika bertransaksi, maka peneliti akan melakukan segmentasi pelanggan.

Peneliti akan mengunakan data transaksi penjualan milik Authen Café & Space untuk dilakukan *data mining*. Peneliti akan menemukan pola pelanggan dari Authen Café & Space yang selanjutnya akan dibuat strategi promosi menggunakan *product bundle*. Untuk membuat *product bundle*, peneliti akan menggabungkan proses *clustering* dan *association rules mining*.

II. KAJIAN TEORI

A. Product bundle

Product bundle merupakan dua atau lebih produk berbeda dikelompokkan menjadi satu paket yang ditawarkan untuk dijual, biasanya dengan harga diskon. Product bundle terbagi menjadi dua jenis, yaitu pure bundling dan mixed bundling. Pure bundling merupakan product bundle yang produknya dijual hanya dalam bentuk kesatuan bundle. Mixed bundling merupakan product bundle yang produknya dijual dalam bentuk bundle dan dijual secara terpisah.

Product bundle dilakukan karena menjadi salah satu strategi pemasaran yang efektif karena memberi nilai tambah bagi penjual dan pembeli. Product bundle dapat memikat pelanggan untuk membeli produk lebih banyak daripada membeli satuan sehingga dapat meningkatkan pendapatan. Product bundle juga dapat meningkatkan persepsi pelanggan terhadap product bundle tersebut, contohnya ketika product bundle menawarkan harga diskon daripada membeli produk secara satuan.

B. Clustering

Clustering atau segmentation merupakan teknik membagi populasi suatu individu kedalam kelompok yang memiliki kemiripan tertentu untuk mendapat informasi mengenai ciri-ciri yang melekat pada kelompok itu dan mengembangkan strategi bisnis yang disesuaikan dengan setiap kelompok [12]. Untuk mengevaluasi seberapa bagus proses clustering yang dilakukan, ada beberapa metrik yang dapat digunakan, seperti: Silhouette index, Davies-bouldin index, Callinzki-harabasz index.

1. Silhouette Index

Silhouette diperkenalkan pertama kali oleh Rousseeuw pada tahun 1987. Silhouette Index mengukur kemiripan *data point* dalam suatu cluster dibandingkan dengan *data point* pada cluster lainnya. Silhouette memiliki rentang nilai dari -1 hingga 1. Apabila Silhouette index bernilai negatif dan semakin mendekati ke -1, maka cenderung dikelompokkan ke cluster yang salah. Apabila Silhouette index bernilai positif dan mendekati ke +1, maka cenderung dikelompokkan ke cluster yang benar [13].

2. Davies-bouldin Index

Davies-bouldin index memberikan peringkat pada cluster yang memiliki sebaran data paling sedikit dengan nilai yang tinggi. Davies-bouldin index berkisar dari 0 hingga 1. Semakin tinggi davies-bouldin index, maka hasil proses clustering semakin buruk. Hal ini menunjukkan bahwa cluster-cluster tidak terpisah dengan baik. Semakin rendah davies-bouldin index, maka hasil proses clustering semakin baik. Hal ini menujukkan bahwa cluster-cluster terpisah dengan baik [14].

3. Calinski-harabaz index

Calinski-harabaz index pertama kali diperkenalkan oleh T. Calinski dan J. Harabasz pada tahun 1974. Metrik ini mengukur derajat penyebaran antara cluster yang satu dengan cluster lainnya. Semakin besar nilai calinski-harabasz indexnya, sehingga semakin baik pula efek pengelompokannya, hal ini disebabkan karena titik-titik data lebih tersebar antar cluster dibandingkan di dalam cluster [15].

C. Association Rules

Association Rules merupakan salah satu teknik data mining yang bertujuan untuk mendapatkan hubungan antara data dalam database yang sering digunakan bersama [9]. Untuk mendapatkan association rules, terdapat measure yang diperlukan, yaitu support dan confidence.

Support merupakan persentase transaksi dalam database yang berisi salah satu atau beberapa item terhadap total transaksi.

$$Supp(P \Rightarrow Q) = Supp(P \cap Q) = \frac{jumlah\ transaksi\ yang\ mengandung\ P\ dan\ Q}{Jumlah\ total\ transaksi} \quad (1)$$

Confidence merupakan persentase transaksi dari salah satu atau beberapa item terhadap jumlah transaksi yang mengandung salah satu atau beberapa item tersebut [10].

$$\text{Conf}(P \Longrightarrow Q) = \frac{\text{Supp}(P \cap Q)}{\text{Supp}(P)} \quad (2)$$

Untuk mengukur kualitas dari rules yang dihasilkan, diperlukan measure berupa lift. Lift merupakan salah satu measure yang berfungsi untuk mengukur seberapa penting rules yang telah dibuat berdasarkan.

$$\text{Lift}(P \Longrightarrow Q) = \frac{Support(P \cap Q)}{Support(P) \times Support(Q)} \quad (3)$$

Jika lift bernilai 1 atau lebih maka berkorelasi positif. Apabila nilai lift kurang dari 1, maka bertolak belakang. jika lift bernilai nol, maka rules tidak ada korelasinya [11].

D. Algoritma Fp-Growth

Algoritma fp-growth merupakan salah satu algoritma untuk association rules dengan cara merancang *frequent pattern tree* (fp-tree) yang merupakan struktur pohon untuk menyimpan informasi pola dari pengembangan prefix-tree. Algoritma fp-growth menggunakan Teknik pencarian *divide and conquer* yang mana akan mengurangi ukuran pola dengan kondisi tertentu (conditional pattern) yang dihasilkan pada tingkat selanjutnya dalam struktur tree[16]. Divide and conquer merupakan teknik membagi data yang ditambang ke dalam subset data berdasarkan identifikasi pola yang sering muncul [17].

E. Algoritma K-Means

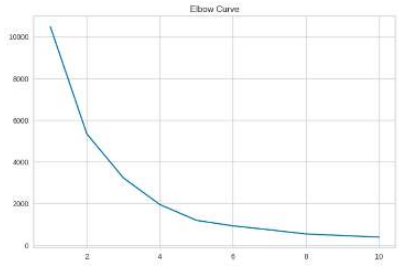
Algoritma k-means merupakan algoritma untuk melakukan *clustering* dengan membagi antar cluster berdasarkan jarak k centroid yang disebut *euclidean distance* [18]. Untuk menentukan jumlah cluster pada k-means dapat menggunakan elbow method. Metode ini bekerja dengan mencari SSE (Sum of Square Error), yaitu jumlah kuadrat jarak antara titik-titik dalam sebuah cluster dengan centroid. Jika nilai cluster sebelumnya dengan nilai cluster saat ini terdapat penurunan dan membentuk sudut, maka nilai cluster tersebut merupakan yang terbaik (Marutho et al., 2018).

III. METODE

Sistematika penyelesaian masalah menggunakan model *Knowledge Discovery in Database (KDD)* yang meliputi 5 tahap, yaitu *selection, preprocessing, transformation, data mininig,* dan *evaluation* [19]. Data yang digunakan merupakan data penjualan Authen Café & Space dari bulan April hingga Agustus 2023. Data didokumentasikan dengan mengunduh langsung melalui aplikasi *point of sales* milik Authen Café & Space.

IV. HASIL DAN PEMBAHASAN

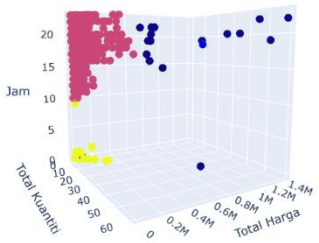
Data yang digunakan untuk menentukan segmentasi pelanggan yaitu total kuantiti *item* yang dibeli, jam kedatangan pelanggan ke *coffee shop*, dan total pengeluaran yang dilakukan pelanggan, sedangkan data yang digunakan untuk melakukan *association rules mining* menggunakan daftar transaksi *item* pembelian oleh pelanggan. Dengan menggunakan visualisasi *elbow method* pada plot WCSS menunjukkan jumlah *cluster* efektif sebanyak 3 cluster.



GAMBAR 1
(Visualisasi *Elbow Method*)

TABEL 1
(Tabel *Centroid* setiap *Cluster*)

Cluster	Total Kuantiti	Total Harga	Jam Kedatangan
0	38,254	926832,407	18,436
1	3,695	88110,666	19,101
2	3,152	71812,828	0,308



GAMBAR 2
(Visualisasi *Cluster* setiap Dataset)

Dari visualisasi grafik diatas maka didapat ciri-ciri setiap cluster. Pada cluster 0, pelanggan mengunjungi coffee shop di rentang pukul 3 sore hingga 12 malam. Pengeluaran rata-rata pelanggan di cluster lebih dari Rp 450.000,00 dengan total item lebih dari 10 item.

Pengunjung pada cluster 1 dominan mengunjungi coffee shop di rentang pukul 10 pagi hingga 11 malam. Rata-rata pengeluaran pelanggan di cluster 1 kurang dari Rp 500.000,00. Kuantiti pembelian item para pelanggan di cluster 1 rata-rata di bawah 10 item pada rentang waktu sebelum pukul 4 sore dan dapat lebih dari 10 item setelah pukul 4 sore.

Pengunjung pada cluster 2 dominan mengunjungi coffee shop di rentang pukul 12 malam hingga pukul 1 malam, Dimana merupakan waktu sebelum coffee shop tutup. Rata-rata pengeluaran pelanggan di cluster 1 lebih dari Rp 200.000,00. Rata-rata pembelian item para pelanggan di cluster 2 tidak lebih dari 10 item.

TABEL 2
(Jumlah *Cluster* dengan *Silhouette Indexnya*)

Jumlah Cluster	Silhouette Index
2	0.740
3	0.723
4	0.613
5	0.518
6	0.543
7	0.542
8	0.506

Hasil pengujian *clustering* menggunakan *silhouette index* ini sejalan dengan teori bahwa semakin tinggi silhouette index mendekati nilai 1, maka akan semakin baik pengelompokkannya.

TABEL 3
(Jumlah *Cluster* dengan *Davies-Bouldin Indexnya*)

Jumlah Cluster	Davies-Bouldin Index
2	0.587
3	0.492
4	0.525
5	0.556
6	0.593
7	0.457
8	0.510

Visualisasi davies-bouldin index dipilih jumlah cluster efektif sebanyak 2 cluster dengan mempertimbangkan juga pada silhouette index dan calinski-harabasz index.

TABEL 4
(Jumlah Cluster dengan Calinski-Harabasz Indexnya)

Jumlah Cluster	Calinski-Harabasz Index
2	3419,121
3	4131,502
4	5083,622
5	6595,247
6	7044,479
7	7855,536
8	9519,978

Pada visualisasi meggunakan metrik calinski-harabasz index menunjukkan semakin tinggi nilai calinski-harabasz index, maka semakin baik pengelompokkannya karena titik data lebih tersebar antar cluster dibanding dalam cluster. Dipilih 2 cluster karena juga mempertimbangkan silhouette index.

Dari proses *clustering,* dataset dibagi menjadi 3 *cluster*. Pada setiap *cluster* dilakukan *association rules mining* menggunakan algoritma *fp-growth* untuk menentukan *product bundle* dari hasil *rules* setiap *cluster.* Setiap *rules* yang terbentuk harus menghasilkan nilai *lift ratio* lebih dari 1 agar *rules* tersebut valid.

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift
0	(Es Kopi Susu Karamel)	(Es Kopi Susu Aren)	0.8125	0.8125	0.6875	0.846154	1.041420
1	(Es Kopi Susu Aren)	(Es Kopi Susu Karamel)	0.8125	0.8125	0.6875	0.846154	1.041420
3	(Ayam Rica-Rica)	(Ayam Popcorn Dabu-Dabu)	0.7500	0.8125	0.6250	0.833333	1.025641
5	(Ayam Rica-Rica)	(Es Kopi Susu Aren)	0.7500	0.8125	0.6250	0.833333	1.025641
2	(Ayam Popcorn Dabu-Dabu)	(Ayam Rica-Rica)	0.8125	0.7500	0.6250	0.769231	1.025641
4	(Es Kopi Susu Aren)	(Ayam Rica-Rica)	0.8125	0.7500	0.6250	0.769231	1.025641

GAMBAR 3
(*Association Rules* pada *Cluster* 0)

Pada *cluster* 0 dengan nilai *minimum lift* sebesar 1 menghasilkan 6 *rules.* Baik dari *support* dan *confidence* memiliki nilai yang tinggi, yaitu di atas 50%. Dari *rules* tersebut dapat disimpulkan menjadi tiga *product bundle,* yaitu *es kopi susu karamel* dengan *es kopi susu aren, ayam rica-rica* dengan *ayam popcorn dabu-dabu, dan ayam rica-rica* dengan *es kopi susu aren.*

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift
2	(Keju Aroma)	(Es Kopi Susu Karamel)	0.058617	0.277257	0.018171	0.310000	1.118097
1	(Platter)	(Es Kopi Susu Karamel)	0.056272	0.277257	0.015826	0.281250	1.014403
3	(Es Kopi Susu Karamel)	(Keju Aroma)	0.277257	0.058617	0.018171	0.065539	1.118097
0	(Es Kopi Susu Karamel)	(Platter)	0.277257	0.056272	0.015826	0.057082	1.014403

GAMBAR 4
(*Association rules* pada *Cluster 1*)

Pada cluster 1 dengan nilai *minimum lift* sebesar 1 menghasilkan 4 *rules*. Dari *rules* yang terbentuk dengan *minimum supportnya* terlalu kecil dan nilai *confidencenya* tidak diatas 50% menunjukkan asosiasinya jarang terjadi dan lemah, walaupun memiliki nilai *lift* yang menunjukkan bahwa anteseden memiliki asosiasi yang positif dengan konsekuen. *Rules* tersebut dapat menjadi *product bundle*, yaitu *keju aroma* atau *platter* dengan *es kopi susu caramel,* namun asosiasi dari *rules* ini lemah.

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift
7	(Pisang Aroma)	(Es Kopi Susu Karamel)	0.054348	0.228261	0.032609	0.600000	2.628571
5	(Tahu Lada Garam)	(Es Kopi Susu Aren)	0.065217	0.195652	0.032609	0.500000	2.555556
0	(Cokelat)	(Es Kopi Susu Aren)	0.119565	0.195652	0.032609	0.272727	1.393939
2	(Cokelat)	(Es Kopi Susu Karamel)	0.119565	0.228261	0.032609	0.272727	1.194805
1	(Es Kopi Susu Aren)	(Cokelat)	0.195652	0.119565	0.032609	0.166667	1.393939
4	(Es Kopi Susu Aren)	(Tahu Lada Garam)	0.195652	0.065217	0.032609	0.166667	2.555556
3	(Es Kopi Susu Karamel)	(Cokelat)	0.228261	0.119565	0.032609	0.142857	1.194805
6	(Es Kopi Susu Karamel)	(Pisang Aroma)	0.228261	0.054348	0.032609	0.142857	2.628571

GAMBAR 5
(*Association Rules pada Cluster 2*)

Pada cluster 2 dengan nilai *minimum lift* sebesar 1 menghasilkan 8 rules dengan item anteseden dan konsekuennya. Dari *rules* yang terbentuk dengan *minimum supportnya* yang sebesar 3% dan sebagian nilai *confidence* lebih besar atau sama dengan 50% mengindikasikan bahwa asosiasi *rules* tersebut jarang terjadi dan memiliki asosiasi yang kuat, ditambah dengan nilai *lift* yang mengindikasikan bahwa anteseden memiliki asosiasi yang positif dengan konsekuen. Dari rules tersebut dapat diambil menjadi dua product bundle, yaitu *pisang aroma* dengan *es kopi susu karamel* dan *tahu lada garam* dengan *es kopi susu aren*.

Dengan nilai *lift ratio* lebih dari 1, diambil hingga 5 *rules* teratas pada setiap *cluster* yang akan menjadi *product bundle* dan diurutkan dari nilai *confidence* tertinggi. Pada cluster 0 menghasilkan 1 *product bundle* yaitu item bernama keju aroma dan es kopi susu karamel.

V. KESIMPULAN

Segmentasi pada dataset penjualan menggunakan algoritma k-means menghasilkan cluster efektif sebanyak 3 *cluster* dengan mempertimbangkan *silhouette index, davies-bouldin index, dan calinski-harabasz index*. Hasil dari *association rules mining* setiap cluster, pada *cluster* pertama menghasilkan 6 rules, pada cluster kedua menghasilkan 4 *rules*, dan *cluster* ketiga menghasilkan 8 *rules*. Berdasarkan hasil *association rules mining* maka saran *product bundle* yang dapat diterapkan dengan mengambil sesuai nilai metrik yang ditentukan, diantaranya, *support, confidence* dan *lift*. Untuk *cluster* 0 *product bundle* yang dapat diterapkan yaitu *es kopi susu karamel* dengan *es kopi susu aren, ayam rica-rica* dengan *ayam popcorn dabu-dabu, dan ayam rica-rica* dengan *es kopi susu aren*. *Product bundle* yang dapt diterapkan pada *cluster* 2 yaitu *pisang aroma* dengan *es kopi susu karamel* dan *tahu lada garam* dengan *es kopi susu aren*..

REFERENSI

[1] Yana Siregar L and Irwan Padli Nasution M, "Perkembangan Teknologi Informasi Terhadap Peningkatan Bisnis Online," *Hirarki : Jurnal Ilmiah Manajemen dan Bisnis*, vol. 2, no. 1, pp. 71–75, 2020.

[2] E. E. Y. Amriel, W. C. Izaak, and A. R. Asnar, "Study of Food Promo Phenomenon on Online Food Delivery in Surabaya," *JOURNAL OF ECONOMICS, FINANCE AND MANAGEMENT STUDIES*, vol. 5, no. 9, Sep. 2022, doi: 10.47191/jefms/v5-i9-08.

[3] Rakuten Insight, "Most used apps for food delivery orders in Indonesia as of April 2023," https://www.statista.com/statistics/1149349/indonesia-favorite-food-delivery-apps/.

[4] Octarini and Harry Susanto E, "Pendapat Pemilik Usaha-Usaha Kuliner Di Jakarta Terhadap Peran Go-Food Dalam Pengembangan Usahanya," *Jurnal Manajemen Bisnis dan Kewirausahaan*, vol. 2, no. 6, Aug. 2019, doi: 10.24912/jmbk.v2i6.4915.

[5] Gojek, "GoFood Konsisten jadi Andalan Pelanggan dan Buktikan Perannya jadi Barometer Tren Kuliner Masyarakat," https://www.gojek.com/blog/gofood/tren-kuliner/. Accessed: Aug. 23, 2023. [Online]. Available: https://www.gojek.com/blog/gofood/tren-kuliner/

[6] I. Syukra, A. Hidayat, and M. Z. Fauzi, "Implementation of K-Medoids and FP-Growth Algorithms for Grouping and Product Offering Recommendations," *Indonesian Journal of Artificial Intelligence and Data Mining*, vol. 2, no. 2, p. 107, Nov. 2019, doi: 10.24014/ijaidm.v2i2.8326.

[7] S. Genjang Setyorini, K. Sari, L. Rahma Elita, and S. A. Putri, "Market Basket Analysis with K-Means and FP-Growth Algorithm as Citra Mustika Pandawa Company," *MALCOM: Indonesian Journal of Machine Learning and Computer Science* , vol. 1, pp. 41–46, 2021.

[8] M. Beladev, L. Rokach, and B. Shapira, "Recommender systems for product bundling," *Knowl Based Syst*, vol. 111, pp. 193–206, Nov. 2016, doi: 10.1016/j.knosys.2016.08.013.

[9] T. A. Kumbhare and S. V Chobe, "An Overview of Association Rule Mining Algorithms," *(IJCSIT) International Journal of Computer Science and Information Technologies*, vol. 5, no. 1, 2014.

[10] A. Ait-Mlouk, F. Gharnati, and T. Agouti, "An improved approach for association rule mining using a multi-criteria decision support system: a case study in road safety," *European Transport Research Review*, vol. 9, no. 3, p. 40, Sep. 2017, doi: 10.1007/s12544-017-0257-5.

[11] N. Hussein, A. Alashqur, and B. Sowan, "Using the interestingness measure lift to generate association rules," *Journal of Advanced Computer Science & Technology*, vol. 4, no. 1, p. 156, Apr. 2015, doi: 10.14419/jacst.v4i1.4398.

[12] P. Michaud, "Clustering techniques," *Future Generation Computer Systems*, vol. 13, no. 2–3, pp. 135–147, Nov. 1997, doi: 10.1016/S0167-739X(97)00017-4.

[13] M. Shutaywi and N. N. Kachouie, "Silhouette Analysis for Performance Evaluation in Machine Learning with Applications to Clustering," *Entropy*, vol. 23, no. 6, p. 759, Jun. 2021, doi: 10.3390/e23060759.

[14] A. A. Vergani and E. Binaghi, "A Soft Davies-Bouldin Separation Measure," in *2018 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, IEEE, Jul. 2018, pp. 1–8. doi: 10.1109/FUZZ-IEEE.2018.8491581.

[15] X. Wang and Y. Xu, "An improved index for clustering validation based on Silhouette index and Calinski-Harabasz index," *IOP Conf Ser Mater Sci Eng*, vol. 569, no. 5, p. 052024, Jul. 2019, doi: 10.1088/1757-899X/569/5/052024.

[16] J. Han, J. Pei, and Y. Yin, "Mining frequent patterns without candidate generation," *ACM SIGMOD Record*, vol. 29, no. 2, pp. 1–12, Jun. 2000, doi: 10.1145/335191.335372.

[17] S. Sidhu, U. K. Meena, A. Nawani, H. Gupta, and N. Thakur, "FP Growth Algorithm Implementation," 2014.

[18] D. T. Larose and C. D. Larose, *Discovering Knowledge in Data*. Wiley, 2014. doi: 10.1002/9781118874059.

[19] G. Mariscal, Ó. Marbán, and C. Fernández, "A survey of data mining and knowledge discovery process models and methodologies," *Knowl Eng Rev*, vol. 25, no. 2, pp. 137–166, Jun. 2010, doi: 10.1017/S0269888910000032.