

Implementasi Gabungan Metode Klasterisasi dan Asosiasi untuk Analisis Data Survei Kepuasan Pelanggan

Reynisa Beta Nurlyan¹, Wiyli Yustanti²

^{1,2}Fakultas Teknik, Universitas Negeri Surabaya

¹reynisa.20044@mhs.unesa.ac.id

²wiyliyustanti@unesa.ac.id

Abstrak— Salah satu cara yang umum digunakan untuk mengukur tingkat kepuasan pelanggan, yaitu dengan melalui pelaksanaan survei kepuasan pelanggan. Survei kepuasan pelanggan diadakan secara berkala oleh penyedia layanan, contohnya, yaitu Universitas Negeri Surabaya Survei tersebut terdiri dari 23 instrumen pertanyaan yang menggunakan Skala Likert sebagai skala penilaian, di mana mahasiswa berperan sebagai responden (pelanggan) yang berjumlah sebanyak 29.409 responden. Metode asosiasi akan diterapkan terhadap data untuk menemukan hubungan keterkaitan tersembunyi di baliknya. Namun, dengan jumlah responden penelitian sedemikian banyak, memungkinkan terciptanya sejumlah variasi data mengenai penilaian mahasiswa terhadap pelayanan akademik yang diberikan oleh perguruan tinggi. Oleh karena itu, metode klasterisasi akan diterapkan terlebih dahulu untuk meminimalisir munculnya *rules* yang tidak relevan. Berdasarkan hasil penelitian, *clustered data* dari hasil implementasi algoritma *K-Means*, yang terdiri dari 3 *cluster*, memiliki *silhouette score* tertinggi, yaitu 0,641. Selanjutnya, algoritma *Apriori* akan diterapkan terhadap *clusters* tersebut, sehingga menghasilkan aturan asosiasi yang bermakna. Pertama, Cluster 1 (Sangat Puas), menghasilkan 3 *rules* dengan *rule* terbaiknya yang memiliki nilai *support* sebesar 88% dan *confidence* sebesar 96%. Kedua, Cluster 2 (Puas), menghasilkan 4 *rules* dengan *rule* terbaiknya yang memiliki nilai *support* sebesar 76% dan *confidence* sebesar 96%. Ketiga, Cluster 3 (Kurang Puas), menghasilkan 3 *rules* dengan *rule* terbaiknya yang memiliki nilai *support* sebesar 54% dan *confidence* sebesar 86%.

Kata Kunci— klasterisasi, asosiasi, survei kepuasan pelanggan, layanan, skala Likert

I. PENDAHULUAN

Seiring dengan perkembangan zaman, kebutuhan manusia akan perbaikan kualitas layanan semakin meningkat dari tahun ke tahun. Hal tersebut menyebabkan munculnya tuntutan terhadap perbaikan kualitas layanan agar dilakukan secara berkala, sehingga kepuasan pelanggan dapat terus mengalami peningkatan. Pada dasarnya, kepuasan pelanggan merupakan penilaian pelanggan sehubungan dengan rasa kepuasan mereka terhadap pilihannya dalam pembelian dan penggunaan suatu produk dan layanan tertentu [1]. Penyedia layanan dikatakan berhasil memenuhi kebutuhan sesuai dengan yang diharapkan oleh pelanggan, apabila kepuasan pelanggan telah mampu tercapai. Kepuasan pelanggan dapat dinilai melalui hasil perbandingan antara kinerja yang dirasakan dan ekspektasi yang diharapkan oleh pelanggan. Hal tersebut menunjukkan, bahwa kepuasan pelanggan mempunyai ketergantungan yang erat terhadap kesenjangan antara kinerja dan ekspektasi dari pembelian dan penggunaan suatu produk dan layanan tertentu [2].

Pelaksanaan survei kepuasan pelanggan menjadi salah satu cara yang paling umum digunakan untuk mengukur tingkat kepuasan pelanggan secara berkala untuk mengelola kualitas pelayanannya [3]. Pada umumnya, survei kepuasan pelanggan terdiri dari sejumlah instrumen pernyataan terkait dengan kualitas layanan yang telah diberikan oleh penyedia layanan. Setiap instrumen pernyataan yang tercantum di dalam survei kepuasan pelanggan telah mewakili salah satu di antara lima dimensi kualitas layanan, yaitu *Reliability*, *Responsiveness*, *Assurance*, *Empathy*, dan *Tangibles* [4]. Salah satu di antara sekian banyak penyedia layanan yang secara rutin melaksanakan survei kepuasan pelanggan adalah perguruan tinggi sebagai penyedia layanan akademik.

Biasanya, perguruan tinggi melaksanakan survei kepuasan pelanggan pada setiap awal semester, baik semester ganjil maupun semester genap, sebagaimana yang dilakukan oleh Universitas Negeri Surabaya (UNESA). Hal itu bertujuan agar pelaksanaan kegiatan akademik terevaluasi secara berkala, sehingga penindaklanjutan dapat dilakukan demi kualitas layanan akademik yang lebih baik pada tahun ajaran berikutnya. Dalam survei kepuasan pelanggan ini, mahasiswa berperan penting sebagai responden (pelanggan). Pada setiap tahun ajaran baru, jumlah mahasiswa terus mengalami pertumbuhan yang signifikan. Oleh karena itu, metode asosiasi akan diterapkan terhadap data untuk menemukan aturan asosiasi dari jawaban yang diberikan oleh responden terkait penilaian mereka terhadap kualitas layanan akademik yang diberikan oleh perguruan tinggi.

Berdasarkan penelitian terdahulu terkait implementasi metode asosiasi terhadap data survei bertipe kategoris yang dilakukan oleh Iffan Ferdiansa dkk., implementasi algoritma *Apriori* terhadap data hasil survei kepuasan pelanggan terhadap layanan akademik Universitas Hasyim Asy'ari dengan mahasiswa sebagai responden, yang berjumlah sebanyak 250 responden, menghasilkan 7 aturan asosiasi dengan *rule* terbaiknya yang memiliki nilai *support* sebesar 32% dan *confidence* sebesar 67% [5]. Namun, dengan puluhan ribu jumlah mahasiswa UNESA, memungkinkan terciptanya sejumlah variasi data mengenai penilaian terkait kepuasan mereka terhadap layanan akademik yang diberikan oleh perguruan tinggi. Hal tersebut dapat menyebabkan turunnya kinerja metode asosiasi dalam menambang aturan asosiasi dari data hasil survei kepuasan pelanggan yang memungkinkan munculnya aturan tidak relevan [6].

Dalam menyelesaikan permasalahan ini, metode klasterisasi akan diterapkan terlebih dahulu untuk mengurangi kemungkinan munculnya aturan-aturan tidak relevan dari hasil implementasi metode asosiasi terhadap data yang berpotensi

untuk mengurangi keakuratan dari prediksi [7]. Data hasil survei kepuasan pelanggan akan dikelompokkan ke dalam sejumlah *clusters* terlebih dahulu, sebelum metode asosiasi diterapkan pada *clustered data* tersebut. Berdasarkan penelitian terdahulu terkait implementasi metode klusterisasi terhadap data survei bertipe kategoris yang dilakukan oleh Muhammad Gilang Suryanata dkk., implementasi algoritma *K-Means* terhadap data hasil survei kepuasan pelanggan terhadap layanan akademik MTs.S Nurul Amaliyah dengan para siswa sebagai responden, yang berjumlah sebanyak 80 responden, menghasilkan 3 *clusters*: Cluster 0 untuk kelompok siswa yang merasa kurang puas (6 responden), Cluster 1 untuk kelompok siswa yang merasa cukup puas (22 responden), dan Cluster 2 untuk kelompok siswa yang merasa puas (52 responden) dengan layanan akademik yang diberikan oleh MTs.S Nurul Amaliyah [8].

Selain itu, penelitian serupa yang dilakukan oleh Luffi Ardiansyah dan Subhan Ajiz Awalludin, implementasi algoritma *K-Means* terhadap data hasil survei kepuasan pelanggan terhadap layanan Universitas Muhammadiyah Prof DR. Hamka *Online Learning* (OLU) dengan para mahasiswa sebagai responden, yang berjumlah sebanyak 403 responden, menghasilkan 2 *clusters*: Cluster 1 untuk kelompok mahasiswa yang merasa puas (94 responden) dan Cluster 2 untuk kelompok mahasiswa yang merasa kurang puas (309 responden) dengan layanan akademik yang diberikan oleh Universitas Muhammadiyah Prof DR. Hamka [9].

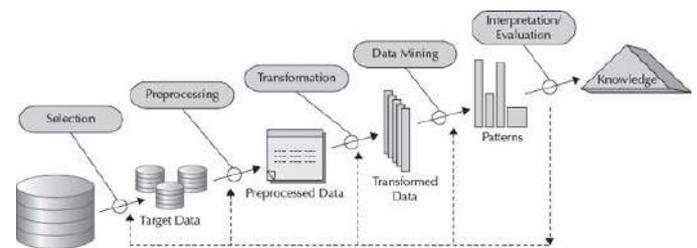
Jadi, pada penelitian ini, kombinasi antara metode klusterisasi dan metode asosiasi akan diterapkan terhadap data hasil survei kepuasan pelanggan dengan menggunakan *library* dari *sklearn* dan *mlxtend*. Pertama-tama, algoritma klusterisasi dengan hasil implementasi terbaik perlu diketahui terlebih dahulu, sehingga perbandingan antara algoritma klusterisasi akan dilakukan untuk menentukan hasil implementasi algoritma klusterisasi terbaik berdasarkan *silhouette*. Dalam mengimplementasikan sejumlah algoritma dari metode klusterisasi terhadap data hasil survei kepuasan pelanggan, *library* dari *sklearn* akan digunakan, yaitu sebagai berikut: *K-Means*, *Agglomerative Clustering*, *Balanced Iterative Reducing and Clustering using Hierarchies* (BIRCH), *Density-Based Spatial Clustering of Applications with Noise* (DBSCAN), *Hierarchical Density-Based Spatial Clustering of Applications with Noise* (HDBSCAN), dan *Spectral Clustering*.

Selanjutnya, *clustered data* dengan *silhouette score* tertinggi akan digunakan untuk penelitian lebih lanjut. Berikutnya, *library* dari *mlxtend* akan digunakan untuk implementasi algoritma *Apriori* dari metode asosiasi terhadap *clustered data* tersebut yang bertujuan untuk mengekstrak *frequent itemsets* dari *clustered data*, di mana *frequent itemsets* tersebut akan digunakan untuk menambang aturan asosiasi yang bermakna dari setiap *cluster*. Hasil penelitian ini diharapkan dapat menjadi referensi bagi perguruan tinggi dalam mengelola kualitas layanan akademiknya.

II. METODE PENELITIAN

Data mining merupakan sebuah proses dalam menemukan pola yang bermakna, menarik, dan baru dari *database* dalam jumlah besar [10]. Terdapat sebagian orang yang menganggap, bahwa istilah *Data Mining* dan *Knowledge Discovery in Database* (KDD) bermakna serupa. Akan tetapi, pada dasarnya, *data mining* adalah bagian dari proses KDD yang meliputi pengaplikasian analisis data dan algoritma penemuan (*discovery algorithms*) untuk menghasilkan enumerasi pola dengan syarat proses KDD masih berada di bawah batasan efisiensi komputasi yang dapat diterima [11]. Proses KDD merupakan proses dalam penggunaan *database*, di mana proses tersebut terdiri dari tahapan seleksi, pra-pemrosesan, subsampling, dan transformasi *database* tersebut [11].

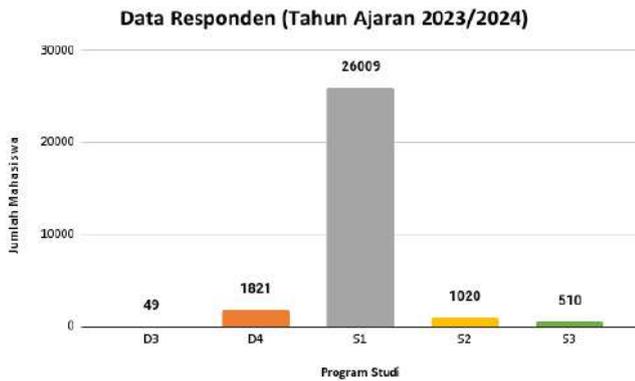
Berikutnya, KDD sendiri didefinisikan sebagai suatu proses non-trivial dalam melakukan identifikasi terhadap pola-pola pada data yang bersifat absah, baru, berpotensi memiliki kegunaan, dan mampu dipahami oleh pengguna [11]. Jadi, dalam penelitian ini metode KDD digunakan untuk menganalisis data hasil survei kepuasan pelanggan secara lebih mendalam. Terdapat sejumlah tahapan yang perlu dilakukan secara berurutan untuk menganalisis data dengan menggunakan metode KDD [11], yang ditunjukkan melalui Gambar 1, yaitu sebagai berikut:



Gbr. 1 Proses KDD

1. Selection

Pada tahap ini, dataset penelitian, yang akan digunakan sebagai data target, perlu ditentukan terlebih dahulu. *Dataset* yang digunakan dalam penelitian ini merupakan data hasil survei kepuasan pelanggan terhadap layanan akademik yang diberikan oleh Universitas Negeri Surabaya pada tahun ajaran 2023/2024 yang, dengan para mahasiswa UNESA sebagai responden, berjumlah sebanyak 29.409 responden. Gambar 2 adalah diagram batang yang menunjukkan jumlah responden dari masing-masing program pendidikan, yaitu program pendidikan Diploma 3 (D3), Diploma 4 (D4), Sarjana 1 (S1), Magister (S2), dan Doktor (S3).



Gbr. 2 Data Jumlah Responden

Jenis dari *dataset* ini berdasarkan sumbernya adalah data sekunder, sebab *dataset* diperoleh melalui observasi yang dilakukan terhadap *web application* resmi dari UNESA, yaitu *Single Sign On UNESA*. Populasi penelitian merupakan seluruh mahasiswa UNESA, sedangkan sampel penelitian merupakan mahasiswa aktif UNESA pada semester ganjil tahun ajaran 2023/2024. Survei kepuasan pelanggan, yang dilakukan oleh UNESA memiliki 23 instrumen pernyataan, yang berperan sebagai variabel penelitian. Berikut adalah rincian dari variabel penelitian beserta definisi operasionalnya yang ditunjukkan melalui Tabel 1, yaitu sebagai berikut:

Tabel I
Variabel Penelitian

Variabel Penelitian	Definisi Operasional
P1	Ketersediaan layanan akademik, administrasi, dan layanan kebutuhan informasi akademis dan nonakademis secara <i>online</i> dan <i>offline</i> dengan akurat dan memuaskan.
P2	Kejelasan informasi program bidang penalaran, minat, dan bakat.
P3	Kemampuan dan kecepatan petugas dalam memberikan layanan bidang penalaran, minat, dan bakat.
P4	Kepedulian petugas dalam menerima keluhan terkait pelayanan bidang penalaran, minat, dan bakat.
P5	Ketersediaan, kecukupan, aksesibilitas, dan kualitas sarana prasarana layanan bidang penalaran, minat, dan bakat.
P6	Kejelasan SOP layanan Bimbingan Konseling, kesehatan, dan beasiswa.
P7	Kemampuan dan kecepatan petugas dalam memberikan layanan Bimbingan Konseling, kesehatan, dan beasiswa.
P8	Kepedulian petugas dalam menerima keluhan terkait layanan Bimbingan Konseling, kesehatan, dan beasiswa.
P9	Ketersediaan, kecukupan, aksesibilitas, dan kualitas sarana prasarana layanan BK, kesehatan, dan beasiswa.
P10	Kejelasan SOP layanan bimbingan karir dan kewirausahaan.

P11	Kemampuan dan kecepatan petugas dalam memberikan layanan bimbingan karir dan kewirausahaan.
P12	Kepedulian petugas dalam menerima keluhan terkait layanan bimbingan karir dan kewirausahaan.
P13	Ketersediaan, kecukupan, aksesibilitas, dan kualitas sarana prasarana layanan bimbingan karir dan kewirausahaan.
P14	Kejelasan SOP untuk proses pengajuan/pembayaran/keterlambatan/keringanan UKT
P15	Kemudahan proses pengajuan/ pembayaran/ keterlambatan/ keringanan UKT.
P16	Kemampuan dan kecepatan petugas dalam memberikan pelayanan proses peminjaman/ penggunaan fasilitas sarana prasarana untuk kegiatan mahasiswa
P17	Kepedulian petugas dalam menerima keluhan terkait layanan keuangan dan sarana prasarana.
P18	Ketersediaan dan kecukupan sarana dan prasarana akademis (perpustakaan, pembelajaran/ bengkel/ laboratorium/ instalasi listrik/ internet, sistem informasi).
P19	Kemampuan dosen, tenaga kependidikan, dan pengelola dalam memberikan pelayanan.
P20	Kemampuan dari dosen, tenaga kependidikan, dan pengelola dalam membantu mahasiswa dan memberikan jasa dengan cepat.
P21	Kemampuan dosen, tenaga kependidikan, dan pengelola untuk memberi keyakinan kepada mahasiswa bahwa pelayanan yang diberikan telah sesuai dengan ketentuan.
P22	Kesediaan/kepedulian pengelola, dosen, tenaga kependidikan untuk memberi perhatian kepada mahasiswa.
P23	Kecukupan, aksesibilitas, kualitas sarana dan prasarana

Dalam pelaksanaan survei kepuasan pelanggan ini, responden perlu mengisi survei dengan berdasarkan skala Likert. Berikut adalah skala Likert yang ditunjukkan melalui Tabel 2, yaitu sebagai berikut:

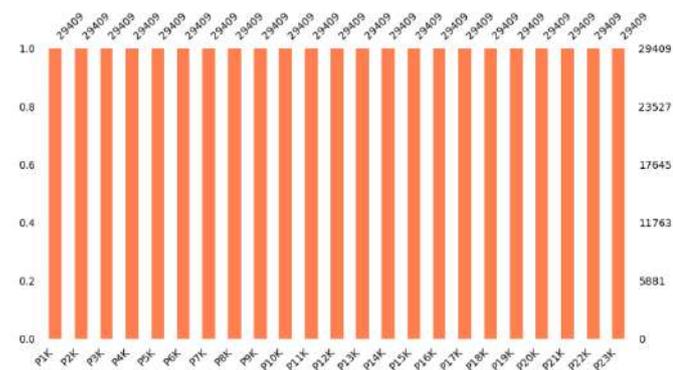
Tabel II
Skala Likert

Tingkat Kinerja Layanan	Nilai
Kurang Baik	1
Cukup Baik	2
Baik	3
Sangat Baik	4

2. Preprocessing

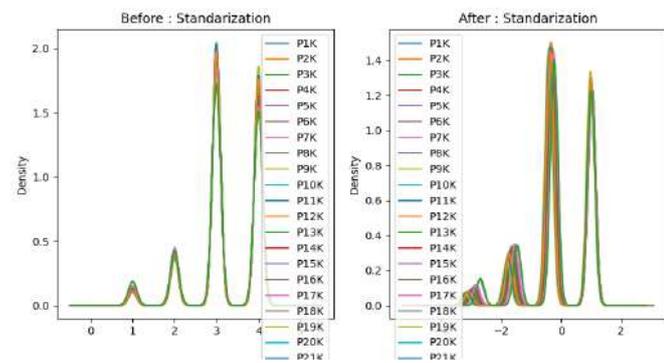
Pada tahap ini, pembersihan data dan pra-pemrosesan data akan diterapkan terhadap *dataset* sebelum *dataset* siap untuk digunakan dalam tahap transformasi data. Pembersihan data dapat berupa penanganan data duplikat dan data hilang. Pertama, untuk menangani data duplikat, setiap responden hanya dapat mengirimkan satu jawaban saja melalui akun kampus mereka, sehingga *dataset* tidak mengandung data duplikat di dalamnya. Kedua, untuk penanganan data hilang,

pemeriksaan dapat dilakukan terhadap *dataset* untuk memastikan tidak ada data hilang terkandung di dalamnya.



Gbr. 3 Pemeriksaan Data Hilang

Berdasarkan hasil pemeriksaan yang ditunjukkan melalui Gambar 3, dapat terlihat, bahwa tidak satu pun variabel penelitian yang mengandung *missing values*. Selanjutnya, pra-pemrosesan data dapat berupa normalisasi data yang bertujuan untuk meminimalisir adanya redundansi di dalam dataset. Hasil normalisasi data ditunjukkan melalui Gambar 4, yaitu sebagai berikut:



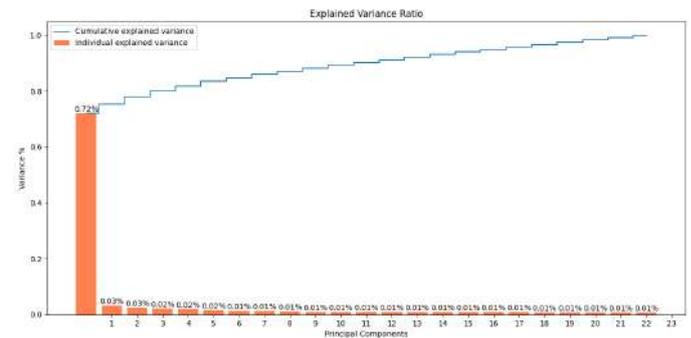
Gbr. 4 Normalisasi Data

Algoritma dari metode klusterisasi dan metode asosiasi merupakan metode yang tergolong ke dalam *unsupervised learning algorithms* [12]. Oleh karena itu, serangkaian tahap pre-processing yang umumnya dilakukan pada *supervised learning algorithms* tidak dilakukan karena *dataset* tidak memiliki atribut *class* atau label.

3. Transformation

Pada tahap ini, reduksi dimensi dataset akan dilakukan untuk menemukan fitur-fitur yang mampu merepresentasikan dataset. *Principal Component Analysis (PCA)* merupakan analisis yang digunakan untuk mereduksi dimensi dataset dari tinggi ke rendah [13]. Meskipun begitu, dimensi yang tersisa dari hasil reduksi tetap mampu mempertahankan keseluruhan informasi di dalam *dataset*. PCA bekerja dengan mengidentifikasi arah dimensi yang mampu menjelaskan varians terbanyak, sebelum data diproyeksikan ke arah tersebut. Dalam implementasi PCA, *explained variance* diperlukan untuk mengukur varians yang dijelaskan oleh setiap

komponen utama, dimana komponen utama pertama dengan varians terbanyak dianggap sebagai yang paling penting.

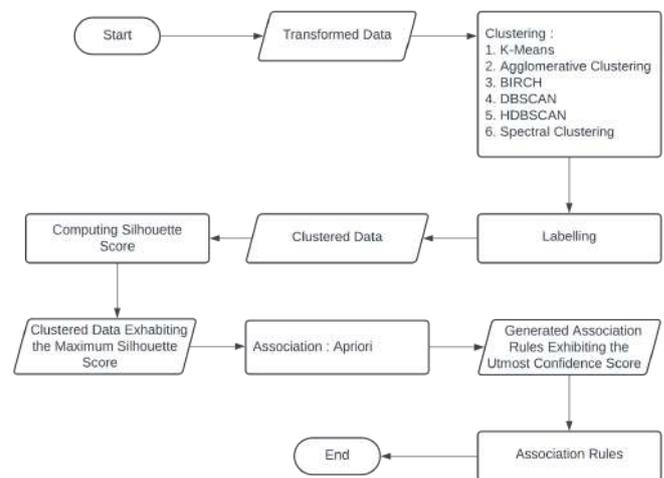


Gbr. 5 Explained Variance Ratio

Berdasarkan hasil komputasi *explained variance* yang ditunjukkan melalui Gambar 5, dapat terlihat bahwa dataset memiliki 32 komponen utama, di mana komponen utama pertama telah mampu menjelaskan 72% varians di dalam dataset. Oleh karena itu, hanya 4 komponen utama saja yang akan digunakan dalam implementasi PCA dalam penelitian ini karena *cumulative explained variance* dari keempat komponen utama tersebut telah mampu menjelaskan >80% varians di dalam *dataset*.

4. Data Mining

Pada tahap ini, algoritma dari metode *data mining* akan diimplementasikan terhadap *transformed data* dengan menyesuaikan *tile goals* dari proses KDD. Diagram alir dari tahap data mining ditunjukkan melalui Gambar 6, yaitu sebagai berikut:



Gbr. 6 Diagram Alir Tahap *Data Mining*

Berdasarkan gambar 6, dapat terlihat, pertama-tama metode klusterisasi akan diterapkan terhadap *transformed data*. Pada dasarnya, klusterisasi merupakan sebuah metode yang digunakan untuk mengelompokkan seluruh *data points* yang berada di dalam suatu *datasets* menuju sejumlah *clusters* berdasarkan kesamaan karakteristiknya [14]. Hal itu

menyebabkan, *data points* yang berada dalam satu *cluster* memiliki tingkat kesamaan yang cukup tinggi antara satu dengan lainnya, tetapi masing-masing *cluster* yang dihasilkan dari proses klusterisasi tersebut memiliki karakteristik yang berbeda. Dalam penelitian ini, *library* dari *sklearn* akan digunakan untuk implementasi algoritma klusterisasi, di mana *library* ini cukup umum digunakan untuk menangani masalah *supervised* dan *unsupervised* dalam skala menengah dengan mengintegrasikan berbagai algoritma *machine learning* [15].

Berdasarkan pendekatannya, metode klusterisasi dibagi menjadi dua secara prinsip, yaitu sebagai berikut: *hierarchical* dan *non-hierarchical clustering* [16]. Pertama, untuk pendekatan *hierarchical clustering*, algoritma klusterisasi yang akan digunakan dalam penelitian ini, adalah sebagai berikut: *K-Means* dan *Spectral Clustering*. Kedua, untuk pendekatan *non-hierarchical clustering*, algoritma klusterisasi yang akan digunakan dalam penelitian ini, adalah sebagai berikut: *Agglomerative Clustering*, *Balanced Iterative Reducing and Clustering using Hierarchies (BIRCH)*, *Density Based Spatial Clustering of Applications with Noise (DBSCAN)*, dan *Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN)*. Model dari setiap algoritma telah diatur parameternya, sehingga proses klusterisasi dari setiap algoritma hanya akan menghasilkan 3 *clusters* saja. Hal tersebut dilakukan untuk mempermudah proses komparasi hasil implementasi algoritma klusterisasi. Pseudocode untuk implementasi masing-masing algoritma dari metode klusterisasi terhadap *transformed data* ditunjukkan melalui Tabel 3, yaitu sebagai berikut:

Tabel III

Pseudocode Implementasi Metode Klusterisasi

Pseudocode : Clustering Methods	
INPUT:	
$D_i = \{D_1, D_2, D_3, \dots, D_n\}$, set of transformed data	
k , number of clusters	
OUTPUT:	
$C_i = \{C_1, C_2, C_3, \dots, C_n\}$, set of clusters	
PROCESS:	
1.	Creating a model for each algorithm of clustering methods (KMeans, Spectral Clustering, BIRCH, DBSCAN, HDBSCAN, and Agglomerative Clustering)
2.	Fitting the said model to the D_i , while simultaneously predicting the cluster label for each data point
3.	Each data point itself is assigned to one of the k clusters based on the algorithm chosen by the user atm

Silhouette akan digunakan dalam proses komparasi hasil implementasi algoritma klusterisasi, sebab *silhouette* merupakan ukuran untuk mengevaluasi validitas proses klusterisasi [17]. *Silhouette* memiliki interval nilai yang berkisar mulai dari nilai -1 hingga 1. Apabila hasil implementasi algoritma klusterisasi memiliki *silhouette score* semakin mendekati -1, maka *data points* dapat dikatakan mengalami *misclustered*. Namun, apabila hasil implementasi algoritma klusterisasi memiliki *silhouette score* semakin mendekati 1, maka dapat dikatakan proses klusterisasi memiliki

validitas yang tinggi. *Clustered data* yang dihasilkan dari implementasi algoritma klusterisasi tertinggi akan digunakan untuk tahapan dari tahap *data mining* yang berikutnya, yaitu implementasi metode asosiasi.

Berikutnya, metode asosiasi akan diterapkan terhadap *clustered data*. Asosiasi merupakan sebuah metode yang digunakan untuk menemukan hubungan atau keterkaitan yang tersembunyi di balik *dataset* [18]. Hubungan keterkaitan tersebut dapat ditampilkan dalam melalui suatu aturan asosiasi. Aturan asosiasi dapat terlihat sebagai implikasi dari bentuk $X \rightarrow Y, X$ sebagai anteseden aturan dan Y sebagai konsekuensi aturan, di mana X dan Y merupakan *item* atau *itemsets* yang berbeda. Dalam penelitian ini, *library* dari *mlxtend* akan digunakan untuk implementasi algoritma asosiasi, di mana *library* tersebut digunakan untuk menyelesaikan permasalahan terkait *machine learning* dan *data mining* dengan mengintegrasikan utilitas dan algoritma yang dikemas menjadi sejumlah *tools* [19].

Dalam penelitian ini, algoritma asosiasi yang akan digunakan dalam penelitian adalah algoritma *Apriori*. Oleh karena itu, algoritma *Apriori* akan diimplementasikan terhadap setiap *cluster* dari *clustered data* terpilih, sehingga tiap *cluster* akan menghasilkan sejumlah *frequent itemsets*, dimana aturan asosiasi yang bermakna akan ditambah dari *frequent itemsets* yang diekstrak dari masing-masing *cluster* tersebut. Namun, sebelum implementasi algoritma *Apriori* diterapkan terhadap *clustered data*, format datanya perlu disesuaikan terlebih dahulu, sehingga berbentuk seperti transaksi penjualan. Dalam proses penyesuaiannya, dapat diketahui, bahwa setiap *cluster* memiliki 23 variabel penelitian. Variabel penelitian ($P_i, i = 1, 2, 3, \dots, 23$) akan dipasangkan dengan skala Likert ($k = 1, 2, 3, 4$) yang hasilnya akan diibaratkan sebagai item produk, yang dapat dituliskan sebagai (P_i_k). Jadi, masing-masing *cluster* tersebut memiliki 92 item produk.

Dalam perumpamaannya apabila item produk terkait 'terbeli', maka kolomnya dapat diisi dengan 'nama item produk' itu sendiri. Namun, apabila item produk tersebut 'tidak terbeli', maka kolomnya boleh diisi dengan angka '0' atau dikosongi. Dalam mempermudah proses penyesuaiannya, fungsi kondisional IF digunakan pada setiap item produk untuk masing-masing *cluster*. Contoh dari format data yang telah disesuaikan untuk implementasi algoritma *Apriori* ditunjukkan melalui Tabel 4, yaitu sebagai berikut:

Tabel IV

Format Data untuk Implementasi Algoritma Apriori						
	P1_1	P1_2	P1_3	...	P23_3	P23_4
0	0	0	0	:	P23_3	0
1	0	0	P1_3	:	0	P23_4
:	:	:	:	:	:	:
n	0	0	0	...	P23_3	0

Pseudocode untuk implementasi algoritma dari metode asosiasi terhadap masing-masing *cluster* ditunjukkan melalui Tabel 5, yaitu sebagai berikut:

Tabel V

Pseudocode Implementasi Metode Asosiasi

Pseudocode : Association Methods

INPUT :

$D_i = \{D_1, D_2, D_3, \dots, D_n\}$, set of clustered data

OUTPUT :

$R_i = \{R_1, R_2, R_3, \dots, R_n\}$, set of association rules

PROCESS :

1. Converting D_i into a list of lists (L_i)
2. Each inner list of L_i will get iterated through each element containing 'item' in the inner list and only keeps the 'item'. if it is not equal to '0', so it would remove occurrences of '0' from the inner list
3. Fitting the encoder to each data point in L_i , then transforming L_i into a binary matrix form
4. Creating a model for Apriori algorithm with a predefined minimum score of support to extract frequent itemsets
5. Creating a model for Association Rules with a predefined minimum score of confidence to generate meaningful rules from frequent itemsets

5. Interpretation/Evaluation

Pada tahap ini, dilakukan interpretasi dan evaluasi terhadap data hasil yang telah diperoleh dari tahap *data mining*. Tahap ini melibatkan visualisasi data untuk setiap hasil implementasi model algoritma, baik dari metode klasterisasi maupun asosiasi, guna memudahkan interpretasi data. Setelah itu, dalam mengevaluasi data, setiap metode memiliki cara tersendiri untuk mengevaluasi kinerjanya. Pertama, untuk metode klasterisasi, *Silhouette* digunakan untuk mengukur kualitas hasil implementasi metode klasterisasi. Apabila nilai *silhouette* semakin mendekati 1 menunjukkan, bahwa validitas proses klasterisasi tergolong tinggi. Kedua, untuk metode asosiasi, *Support-Confidence Framework* digunakan untuk mengukur ketertarikan hasil implementasi metode asosiasi secara objektif. Apabila nilai *support* dan *confidence* semakin mendekati 100% menunjukkan, bahwa aturan asosiasi yang dihasilkan semakin baik akurasi [20].

III. HASIL DAN PEMBAHASAN

A. Hasil Analisis Klasterisasi

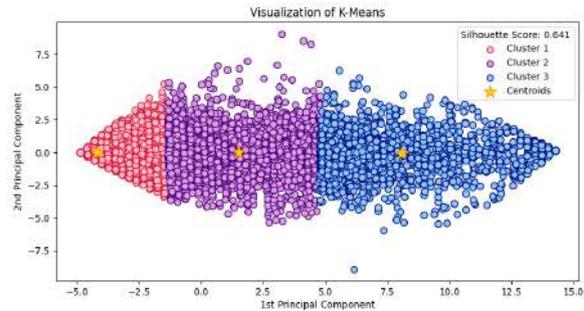
Pada sub-bab ini, implementasi sejumlah algoritma dari metode klasterisasi telah dilakukan terhadap *transformed data*. Berikut adalah visualisasi sekaligus penjelasannya dari hasil implementasi untuk setiap algoritma klasterisasi, yaitu:

1. K-Means

Berdasarkan visualisasi hasil implementasi algoritma *K-Means* terhadap *transformed data* yang ditunjukkan melalui Gambar 7, dapat terlihat, bahwa *transformed data* dikelompokkan menjadi 3 *clusters*, di mana masing-masing cluster memiliki *centroid*, dengan rincian sebagai berikut:

- a. Cluster 1, ditandai dengan *marker* 'point' yang berwarna 'pink', memiliki 11.373 *data points*

- b. Cluster 2, ditandai dengan *marker* 'point' yang berwarna 'plum', memiliki 14.919 *data points*
- c. Cluster 3, ditandai dengan *marker* 'point' yang berwarna 'lightskyblue', memiliki 3.117 *data points*
- d. Centroid, ditandai dengan *marker* '*'

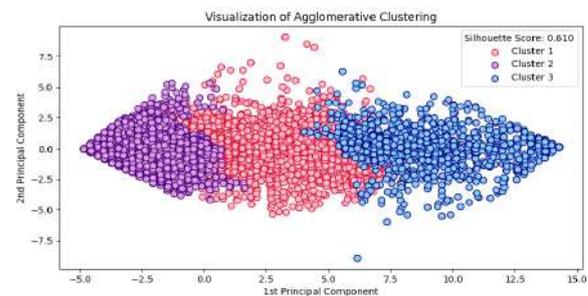


Gbr. 7 Visualisasi Algoritma K-Means

2. Agglomerative Clustering

Berdasarkan visualisasi hasil implementasi algoritma *Agglomerative Clustering* terhadap *transformed data* yang ditunjukkan melalui Gambar 8, dapat terlihat, bahwa *transformed data* dikelompokkan menjadi 3 *clusters*. Namun, dalam implementasi algoritma klasterisasi ini, *Ward's method* dipilih sebagai *linkage criteria*, sebab *linkage criteria* tersebut memiliki *silhouette score* tertinggi dibandingkan dengan *average linkage*, *single linkage*, dan *complete linkage*. Rincian untuk masing-masing *cluster*, yaitu sebagai berikut:

- a. Cluster 1, ditandai dengan *marker* 'point' yang berwarna 'pink', memiliki 14.704 *data points*
- b. Cluster 2, ditandai dengan *marker* 'point' yang berwarna 'plum', memiliki 12.549 *data points*
- c. Cluster 3, ditandai dengan *marker* 'point' yang berwarna 'lightskyblue', memiliki 2.156 *data points*



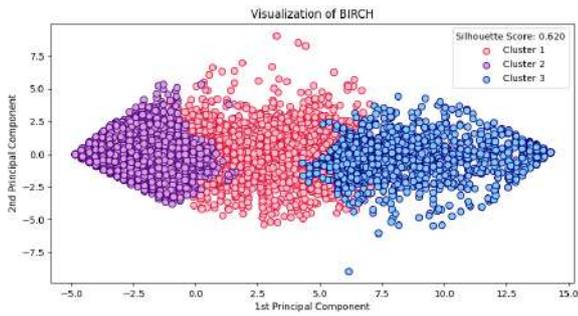
Gbr. 8 Visualisasi Algoritma Agglomerative Clustering

3. BIRCH

Berdasarkan visualisasi hasil implementasi algoritma *BIRCH* terhadap *transformed data* yang ditunjukkan melalui Gambar 9, dapat terlihat, bahwa *transformed data* dikelompokkan menjadi 3 *clusters*, dengan rincian sebagai berikut:

- a. Cluster 1, ditandai dengan *marker* 'point' yang berwarna 'pink', memiliki 14.634 *data points*
- b. Cluster 2, ditandai dengan *marker* 'point' yang berwarna 'plum', memiliki 12.341 *data points*

- c. Cluster 3, ditandai dengan *marker* 'point' yang berwarna 'lightskyblue', memiliki 2.434 *data points*

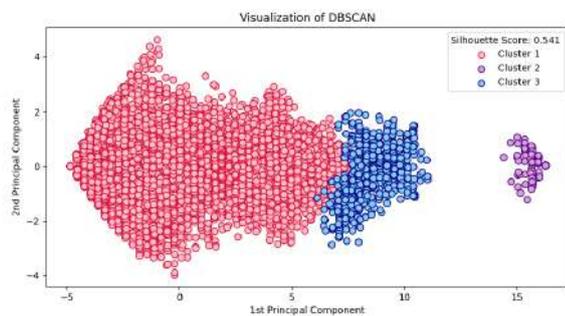


Gbr. 9 Visualisasi Algoritma BIRCH

4. DBSCAN

Berdasarkan visualisasi hasil implementasi algoritma DBSCAN terhadap *transformed data* yang ditunjukkan melalui Gambar 10, dapat terlihat, bahwa *transformed data* dikelompokkan menjadi 3 *clusters*. Namun, dalam pngimplementasian algoritma ini, sebanyak 2.660 *data points* diklasifikasikan sebagai *noise*, sehingga *noise* tersebut perlu ditangani terlebih dahulu agar tidak mengurangi validitas proses klusterisasi. Dari penanganan *noise* tersebut, jumlah *data points* yang tersisa hanya sebanyak 26.749 *data points* saja secara keseluruhan. Rincian untuk masing-masing *cluster*, yaitu sebagai berikut:

- a. Cluster 1, ditandai dengan *marker* 'point' yang berwarna 'pink', memiliki 25.171 *data points*
- b. Cluster 2, ditandai dengan *marker* 'point' yang berwarna 'plum', memiliki 278 *data points*
- c. Cluster 3, ditandai dengan *marker* 'point' yang berwarna 'lightskyblue', memiliki 1.300 *data points*

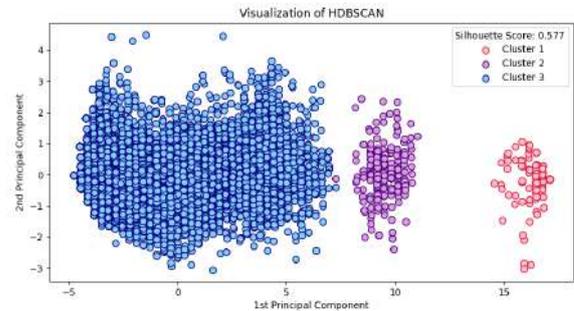


Gbr. 10 Visualisasi Algoritma DBSCAN

5. HDBSCAN

Berdasarkan visualisasi hasil implementasi algoritma HDBSCAN terhadap *transformed data* yang ditunjukkan melalui Gambar 11, dapat terlihat, bahwa *transformed data* dikelompokkan menjadi 3 *clusters*. Namun, dalam pngimplementasian algoritma ini, sebanyak 5.380 *data points* diklasifikasikan sebagai *noise*, sehingga *noise* tersebut perlu ditangani terlebih dahulu agar tidak mengurangi validitas proses klusterisasi. Dari penanganan *noise* tersebut, jumlah *data points* yang tersisa hanya sebanyak 24.029 *data points* saja secara keseluruhan. Rincian untuk masing-masing *cluster*, yaitu sebagai berikut:

- a. Cluster 1, ditandai dengan *marker* 'point' yang berwarna 'pink', memiliki 295 *data points*
- b. Cluster 2, ditandai dengan *marker* 'point' yang berwarna 'plum', memiliki 914 *data point*
- c. Cluster 3, ditandai dengan *marker* 'point' yang berwarna 'lightskyblue', memiliki 23.107 *data points*

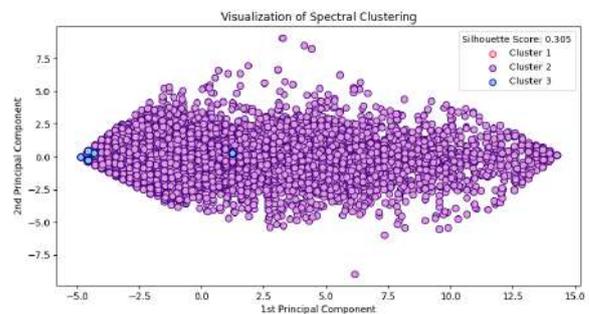


Gbr. 11 Visualisasi Algoritma HDBSCAN

6. Spectral Clustering

Berdasarkan visualisasi hasil implementasi algoritma Spectral Clustering terhadap *transformed data* yang ditunjukkan melalui Gambar 12, dapat terlihat, bahwa *transformed data* dikelompokkan menjadi 3 *clusters*, dengan rincian sebagai berikut:

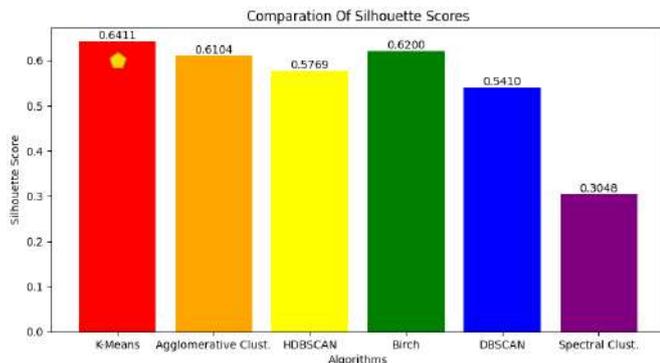
- a. Cluster 1, ditandai dengan *marker* 'point' yang berwarna 'pink', memiliki 101 *data points*
- b. Cluster 2, ditandai dengan *marker* 'point' yang berwarna 'plum', memiliki 22.593 *data points*
- c. Cluster 3, ditandai dengan *marker* 'point' yang berwarna 'lightskyblue', memiliki 6.715 *data points*



Gbr. 12 Visualisasi Algoritma Spectral Clustering

Meskipun legenda untuk masing-masing visualisasi algoritma klusterisasi telah menunjukkan *silhouette score* dari hasil implementasi algoritma tersebut. Visualisasi tersendiri untuk *silhouette score* akan ditampilkan dalam bentuk diagram batang, sebagaimana yang ditunjukkan melalui Gambar 12. Hal tersebut bertujuan untuk membantu pengguna dalam membaca hasil komparasi antara algoritma klusterisasi, sehingga interpretasi data dapat dilakukan dengan mudah. Setiap batang dari diagram tersebut memuat memuat informasi mengenai *silhouette score* untuk masing-masing implementasi algoritma klusterisasi. Berdasarkan Gambar 13, dapat diketahui sebagai berikut:

1. *K-Means*, implementasi algoritma memiliki *silhouette score* sebesar 0,6411
2. *Agglomerative Clustering*, implementasi algoritma memiliki *silhouette score* sebesar 0,6104
3. *BIRCH*, implementasi algoritma memiliki *silhouette score* sebesar 0,5769
4. *DBSCAN*, implementasi algoritma memiliki *silhouette score* sebesar 0,6200
5. *HDBSCAN*, implementasi algoritma memiliki *silhouette score* sebesar 0,5410
6. *Spectral Clustering*, implementasi algoritma memiliki *silhouette score* sebesar 0,3048



Gbr. 13 Komparasi Algoritma Klasterisasi

Berdasarkan Gambar 13 dapat disimpulkan, bahwa, hasil implementasi algoritma *K-Means* terhadap *transformed data* memiliki *silhouette score* tertinggi, yaitu 0.6411. Hal tersebut menunjukkan, bahwa algoritma *K-Means* dapat bekerja dengan baik dalam mengelompokkan data survei bertipe kategoris dengan jumlah yang besar menuju ke *k clusters*. Oleh karena itu, dalam penelitian ini, *clustered data*, yang dihasilkan melalui implementasi algoritma *K-Means*, akan diproses lebih lanjut untuk pengimplementasian algoritma *Apriori*. Namun, sebelumnya, karakteristik responden dari setiap *cluster* perlu dianalisis terlebih dahulu dalam upaya mempermudah pemberian nama secara deskriptif untuk masing-masing *cluster*, di mana pemberian nama tersebut berdasarkan skala Likert. Hasil analisis statistik deskriptif ditunjukkan melalui Tabel 6, yaitu sebagai berikut:

Tabel VI
Analisis Statistik Deskriptif

Variabel Penelitian	\bar{X} Cluster 1	\bar{X} Cluster 2	\bar{X} Cluster 3
P1	3.894	3.038	2.044
P2	3.898	3.026	2.059
P3	3.890	3.016	2.020
P4	3.890	2.997	1.957
P5	3.888	2.975	1.897
P6	3.891	2.985	1.916
P7	3.896	2.996	1.923
P8	3.896	2.991	1.923
P9	3.894	2.988	1.880
P10	3.901	3.010	2.000
P11	3.907	3.010	1.970
P12	3.901	3.004	1.973
P13	3.902	2.997	1.929

P14	3.887	2.989	1.972
P15	3.877	2.978	1.929
P16	3.890	2.988	1.934
P17	3.877	2.959	1.856
P18	3.872	2.947	1.836
P19	3.921	3.066	2.133
P20	3.904	3.042	1.118
P21	3.904	3.042	2.098
P22	3.907	3.030	2.072
P23	3.861	2.948	1.856
Grand Mean	3.893 (4)	3.001 (3)	1.969 (2)
Label Name	Sangat Puas	Puas	Kurang Puas

Analisis statistik deskriptif dilakukan terhadap *clustered data* dengan menghitung nilai *mean* dari tiap variabel penelitian untuk masing-masing *cluster*. Kemudian, nilai *mean* dari seluruh variabel penelitian akan dirata-rata ulang, sehingga menghasilkan *grand mean* untuk masing-masing *cluster*. Selanjutnya, nilai *grand mean* dari tiap *cluster* akan dibulatkan ke ribuan terdekat, yaitu: Pertama, Cluster 1, dengan *grand mean* sebesar 4 Kedua, Cluster 2, dengan *grand mean* sebesar 3. Ketiga, Cluster 3, dengan *grand mean* sebesar 2. Berdasarkan penilaian skala Likert, maka, hasil analisis statistik deskriptif untuk masing-masing *cluster* dapat disimpulkan sebagai berikut, yaitu sebagai berikut:

1. Cluster 1 diberi label nama “Sangat Puas” sebab *cluster* tersebut terdiri dari 11.373 responden yang merasa **sangat puas** dengan pelayanan akademik yang diberikan oleh perguruan tinggi.
2. Cluster 2 diberi label nama “Puas” sebab *cluster* tersebut terdiri dari 14.919 responden yang merasa **puas** dengan pelayanan akademik yang diberikan oleh perguruan tinggi.
3. Cluster 3 diberi label nama “Kurang Puas” sebab *cluster* tersebut terdiri dari 3.117 responden yang merasa **kurang puas** dengan pelayanan akademik yang diberikan oleh perguruan tinggi.

B. Hasil Analisis Asosiasi

Pada sub-bab ini, implementasi algoritma dari metode asosiasi telah dilakukan terhadap *clustered data* terpilih, yaitu *clustered data* dari hasil implementasi algoritma *K-Means*. Berikut adalah visualisasi sekaligus penjelasannya dari hasil implementasi algoritma asosiasi, yaitu sebagai berikut:

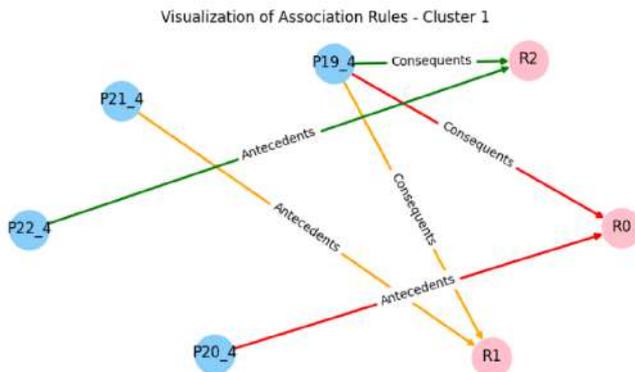
1. Cluster 1

Dalam pengimplementasian algoritma *Apriori* terhadap *clustered data*, *frequent itemset* diekstrak dari *clustered data* tersebut dengan nilai minimum *support* sebesar 87,5%. Setelah itu, aturan asosiasi akan ditambah dari *frequent itemset* yang telah diperoleh sebelumnya dengan menggunakan nilai *threshold* sebesar 95% untuk *metric* ‘confidence’. Hasil penambangan aturan asosiasi untuk Cluster 1 dapat ditunjukkan melalui Tabel 7 yaitu sebagai berikut:

Tabel VII
Aturan Asosiasi Cluster 1

Antecedents	Consequents	Support	Confidence	Lift
(P20_4)	(P19_4)	0.878572	0.959846	1.03
(P21_4)	(P19_4)	0.880419	0.958732	1.03
(P22_4)	(P19_4)	0.875758	0.953110	1.03

Berdasarkan Tabel 7, dapat terlihat, bahwa hasil penambangan aturan terhadap Cluster 1 menghasilkan 3 aturan asosiasi. Ada item produk yang bertindak, baik sebagai anteseden aturan maupun konsekuensi aturan yang dilengkapi dengan nilai *support*, *confidence*, dan *lift*. Visualisasi hasil penambangan aturan terhadap Cluster 1 ditunjukkan melalui Gambar 14, yaitu sebagai berikut:



Gbr. 14 Visualisasi Aturan Asosiasi Cluster 1

Berikut adalah penjelasan untuk masing-masing aturan asosiasi yang ditambang melalui *frequent itemsets* milik Cluster 1, yaitu sebagai berikut:

- $R0 \mid (P20_4) \rightarrow (P19_4)$, aturan pertama: apabila responden memberi nilai 4 pada variabel penelitian P20, maka, kemungkinan besar, responden tersebut juga akan memberikan nilai 4 pada variabel penelitian P19.
- $R1 \mid (P21_4) \rightarrow (P19_4)$, aturan kedua: apabila responden memberi nilai 4 pada variabel penelitian P21, maka, kemungkinan besar, responden tersebut juga akan memberikan nilai 4 pada variabel penelitian P19.
- $R2 \mid (P22_4) \rightarrow (P19_4)$, aturan ketiga: apabila responden memberi nilai 4 pada variabel penelitian P22, maka, kemungkinan besar, responden tersebut juga akan memberikan nilai 4 pada variabel penelitian P19.

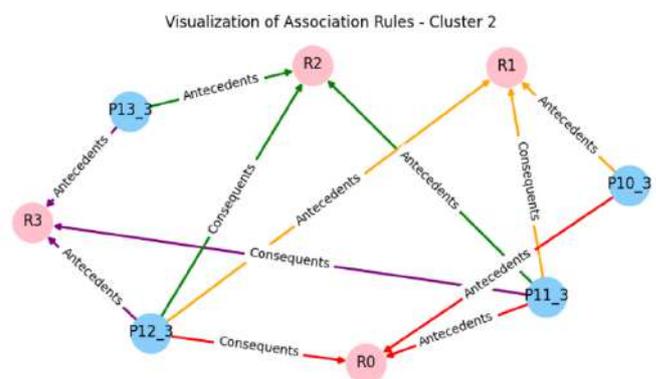
2. Cluster 2

Dalam pengimplementasian algoritma *Apriori* terhadap *clustered data*, *frequent itemset* diekstrak dari *clustered data* tersebut dengan nilai minimum *support* sebesar 75,5%. Setelah itu, aturan asosiasi akan ditambang dari *frequent itemset* yang telah diperoleh sebelumnya dengan menggunakan nilai *threshold* sebesar 95% untuk *metric* 'confidence'. Hasil penambangan aturan asosiasi untuk Cluster 2 dapat ditunjukkan melalui Tabel 8 yaitu sebagai berikut:

Tabel VIII
Aturan Asosiasi Cluster 2

Antecedents	Consequents	Support	Confidence	Lift
(P10_3, P11_3)	(P12_3)	0.758496	0.957603	1.11
(P10_3, P12_3)	(P11_3)	0.758496	0.958577	1.11
(P11_3, P13_3)	(P12_3)	0.759434	0.953110	1.11
(P13_3, P12_3)	(P11_3)	0.759434	0.953543	1.11

Berdasarkan Tabel 8, dapat terlihat, bahwa hasil penambangan aturan terhadap Cluster 2 menghasilkan 4 aturan asosiasi. Ada item produk yang bertindak, baik sebagai anteseden aturan maupun konsekuensi aturan yang dilengkapi dengan nilai *support*, *confidence*, dan *lift*. Visualisasi hasil penambangan aturan terhadap Cluster 2 ditunjukkan melalui Gambar 15 yaitu sebagai berikut:



Gbr. 15 Visualisasi Aturan Asosiasi Cluster 2

Berikut adalah penjelasan untuk masing-masing aturan asosiasi yang ditambang melalui *frequent itemsets* milik Cluster 2, yaitu sebagai berikut:

- $R0 \mid (P10_3, P11_3) \rightarrow (P12_3)$, aturan pertama: apabila responden memberi nilai 3 pada variabel penelitian P10 dan P11, maka, kemungkinan besar, responden tersebut juga akan memberikan nilai 3 pada variabel penelitian P12.
- $R1 \mid (P10_3, P12_3) \rightarrow (P11_3)$, aturan kedua: apabila responden memberi nilai 3 pada variabel penelitian P10 dan P12, maka, kemungkinan besar, responden tersebut juga akan memberikan nilai 3 pada variabel penelitian P11.
- $R2 \mid (P11_3, P13_3) \rightarrow (P12_3)$, aturan ketiga: apabila responden memberi nilai 3 pada variabel penelitian P11 dan P13, maka, kemungkinan besar, responden tersebut juga akan memberikan nilai 3 pada variabel penelitian P12.
- $R3 \mid (P13_3, P12_3) \rightarrow (P11_3)$, aturan keempat: apabila responden memberi nilai 3 pada variabel penelitian P13 dan P12, maka, kemungkinan besar, responden tersebut juga akan memberikan nilai 3 pada variabel penelitian P11.

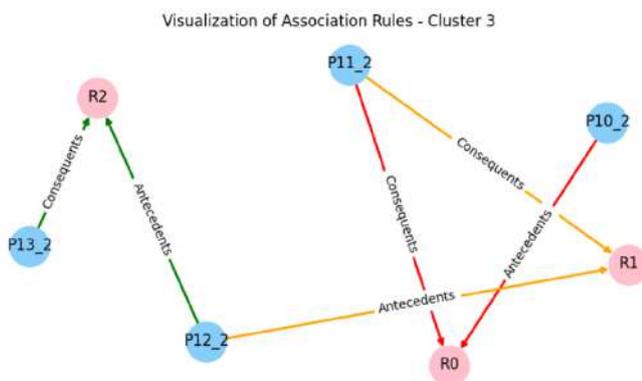
3. Cluster 3

Dalam pengimplementasian algoritma *Apriori* terhadap *clustered data*, *frequent itemset* diekstrak dari *clustered data* tersebut dengan nilai minimum *support* sebesar 50,5%. Setelah itu, aturan asosiasi akan ditambang dari *frequent itemset* yang telah diperoleh sebelumnya dengan menggunakan nilai *threshold* sebesar 80,5% untuk *metric* ‘confidence’. Hasil penambangan aturan asosiasi untuk Cluster 3 dapat ditunjukkan melalui Tabel 9, yaitu sebagai berikut:

Tabel IX
Aturan Asosiasi Cluster 3

Antecedents	Consequents	Support	Confidence	Lift
(P10_2)	(P11_2)	0.526147	0.854167	1.36
(P12_2)	(P11_2)	0.531601	0.851490	1.35
(P12_2)	(P13_2)	0.535451	0.857657	1.35

Berdasarkan Tabel 9, dapat terlihat, bahwa hasil penambangan aturan terhadap Cluster 3 menghasilkan 3 aturan asosiasi. Ada item produk yang bertindak, baik sebagai anteseden aturan maupun konsekuensi aturan yang dilengkapi dengan nilai *support*, *confidence*, dan *lift*. Visualisasi hasil penambangan aturan ditunjukkan melalui Gambar 16, yaitu:



Gbr. 16 Visualisasi Aturan Asosiasi Cluster

Berikut adalah penjelasan untuk masing-masing aturan asosiasi yang ditambang melalui *frequent itemsets* milik Cluster 3, yaitu sebagai berikut:

- $R0 \mid (P10_2) \rightarrow (P11_2)$, aturan pertama: apabila responden memberi nilai 2 pada variabel penelitian P10, maka, kemungkinan besar, responden tersebut juga akan memberikan nilai 2 pada variabel penelitian P11.
- $R1 \mid (P12_2) \rightarrow (P11_2)$, aturan kedua: apabila responden memberi nilai 2 pada variabel penelitian P12, maka, kemungkinan besar, responden tersebut juga akan memberikan nilai 2 pada variabel penelitian P11.
- $R2 \mid (P12_2) \rightarrow (P13_2)$, aturan ketiga: apabila responden memberi nilai 2 pada variabel penelitian P12, maka, kemungkinan besar, responden tersebut juga akan memberikan nilai 2 pada variabel penelitian P13.

IV. KESIMPULAN DAN SARAN

A. Kesimpulan

Menurut hasil komparasi antara implementasi algoritma klusterisasi terhadap dataset, implementasi algoritma *K-Means* memiliki *silhouette score* tertinggi, yaitu 0.641. *Clustered data* yang dihasilkan melalui pengimplementasian *K-Means* terdiri dari 3 cluster, di mana: Cluster 1 merupakan kelompok mahasiswa yang merasa **sangat puas**, Cluster 2 merupakan kelompok mahasiswa yang merasa **puas**, dan Cluster 3 merupakan kelompok mahasiswa yang merasa **kurang puas** dengan pelayanan akademik yang diberikan oleh perguruan tinggi. Masing-masing *cluster* tersebut diimplementasikan dengan algoritma *Apriori*, dengan kesimpulan sebagai berikut:

1. Cluster 1

Aturan asosiasi terbaiknya memiliki nilai *support* sebesar 88% dan *confidence* sebesar 96%, yaitu apabila mahasiswa merasa sangat puas dengan layanan terkait “Kemauan dari dosen, tenaga kependidikan, dan pengelola dalam membantu mahasiswa dan memberikan jasa dengan cepat”, maka, kemungkinan, responden juga merasa sangat puas dengan layanan terkait “Kemampuan dosen, tenaga kependidikan, dan pengelola dalam memberikan pelayanan” atau dapat dituliskan sebagai, (P20_4) → (P19_4).

2. Cluster 2

Aturan asosiasi terbaiknya memiliki nilai *support* sebesar 76% dan *confidence* sebesar 96%, yaitu apabila mahasiswa merasa puas dengan layanan terkait “Kemampuan dan kecepatan petugas dalam memberikan layanan bimbingan karir dan kewirausahaan” dan “Ketersediaan, kecukupan, aksesibilitas, dan kualitas sarana prasarana layanan bimbingan karir dan kewirausahaan”, maka, kemungkinan besar, responden tersebut juga akan merasa puas dengan layanan terkait “Kepedulian petugas dalam menerima keluhan terkait layanan bimbingan karir dan kewirausahaan” atau dapat dituliskan sebagai, (P11_3, P13_3) → (P12_3).

3. Cluster 3

Association rule terbaiknya memiliki nilai *support* sebesar 54% dan *confidence* sebesar 86%, yaitu apabila mahasiswa merasa kurang puas dengan layanan terkait “Kepedulian petugas dalam menerima keluhan terkait layanan bimbingan karir dan kewirausahaan”, maka, kemungkinan, responden juga merasa kurang puas dengan layanan terkait “Ketersediaan, kecukupan, aksesibilitas, dan kualitas sarana prasarana layanan bimbingan karir dan kewirausahaan” atau dapat dituliskan sebagai, (P12_2) → (P13_2).

B. Saran

- Berdasarkan hasil penelitian, diketahui, bahwa algoritma *K-Means* merupakan algoritma klusterisasi dengan hasil implementasi terbaik untuk data survei bertipe kategoris. Oleh karena itu, penggunaan varian dari algoritma *K-Means* dapat dipertimbangkan, seperti *MiniBatch K-Means* dan *Bisecting K-Means*.
- Pertimbangan untuk *Importance-Performance Analysis* (IPA) terhadap variabel penelitian agar tidak hanya

kepuasan pelanggan saja yang diketahui, tetapi juga penting ada atau pun tidaknya layanan bagi pelanggan.

REFERENSI

- [1] G. Guido, "Customer Satisfaction," in *Wiley Encyclopedia of Management*, C. L. Cooper, Ed., Wiley, 2015.
- [2] H. Lee, Y. Lee, and D. Yoo, "The determinants of perceived service quality and its relationship with satisfaction," *J. Serv. Mark.*, vol. 14, no. 3, pp. 217–231, 2000.
- [3] R. S. Kenett and S. Salini, "Modern analysis of customer satisfaction surveys: comparison of models and integrated analysis," *Appl. Stoch. Models Bus. Ind.*, vol. 27, no. 5, pp. 465–475, 2011.
- [4] P. Kotler and K. L. Keller, *Marketing Management*, 14th [ed.]. Upper Saddle River, N.J: Prentice Hall, 2012.
- [5] I. Ferdiansa, A. I. Agung, and A. Faizah, "Implementasi Algoritma Apriori Untuk Sistem Penilaian Kepuasan Mahasiswa Terhadap Pelayanan Di Universitas Hasyim Asy'ari (Studi Kasus di Universitas Hasyim Asy'ari)," *Inov. J. Ilm. Inov. Teknol. Inf.*, vol. 6, no. 2, pp. 1–8, 2022.
- [6] H. Z. Oo and N. S. M. Kham, "Pattern Discovery Using Association Rule Mining on Clustered Data," *Int. J. New Technol. Res.*, vol. 4, no. 2, pp. 8–11, 2018.
- [7] M. Bouraoui and A. Grissa, "Efficient Mining of Association Rules based on Clustering from Distributed Data," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 4, pp. 401–409, 2019.
- [8] Mhd. G. Suryanata, D. H. Pane, and M. Hutasuhut, "Implementasi Algoritma K-Means Dalam Mengukur Tingkat Kepuasan Siswa Terhadap Pelayanan Sekolah Studi Kasus Di Mts. S Nurul Amaliyah," *J. Teknol. Sist. Inf. Dan Sist. Komput. TGD*, vol. 2, no. 2, pp. 118–125, 2019.
- [9] L. Ardiansyah and S. A. Awalludin, "Penerapan Algoritma K-Mean untuk Menentukan Tingkat Kepuasan Mahasiswa Terhadap SISTEM Online Learning UHAMKA (OLU)," *J. Pembelajaran dan Mat. Sigma*, vol. 9, no. 1, pp. 162–171, 2023.
- [10] M. J. Zaki and W. Meira, *Data mining and analysis: fundamental concepts and algorithms*. New York, NY: Cambridge University Press, 2014.
- [11] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "Knowledge Discovery and Data Mining: Towards a Unifying Framework," *Proc. 2nd Int. Conf. Knowl. Discov. Data Min. KDD-96*, pp. 82–88, 1996.
- [12] M. E. Celebi and K. Aydin, Eds., *Unsupervised Learning Algorithms*. Cham: Springer International Publishing, 2016. Accessed: Apr. 22, 2024.
- [13] S. Wold, K. Esbensen, and P. Geladi, "Principal Component Analysis," *Chemom. Intell. Lab. Syst.*, vol. 2, no. 1–3, pp. 37–52, 1987.
- [14] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements Of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd [ed.]. Cham: Springer International Publishing, 2009.
- [15] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python," *J. Mach. Learn. Res.*, pp. 2825–2830, 2011.
- [16] F. Gorunescu, *Data Mining*, vol. 12. in *Intelligent Systems Reference Library*, vol. 12. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011.
- [17] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *J. Comput. Appl. Math.*, vol. 20, pp. 53–65, 1987.
- [18] R. Agrawal and R. Srikant, "Fast Algorithms for Mining Association Rules," *Proc. 20th VLDB Conf.*, 1994.
- [19] S. Raschka, "MLxtend: Providing machine learning and data science utilities and extensions to Python's scientific computing stack," *J. Open Source Softw.*, vol. 3, no. 24, p. 638, Apr. 2018, doi: 10.21105/joss.00638.
- [20] M. Maraghi, M. A. Adibi, and E. Mehdizadeh, "Using RFM Model and Market Basket Analysis for Segmenting Customers and Assigning Marketing Strategies to Resulted Segments," *J. Appl. Intell. Syst. Inf. Sci.*, vol. 1, no. 1, pp. 35–43, 2020.