Jurnal Komputer Multidisipliner

Vol.7 No.10, Oktober 2024
ISSN: 24559633

IMPLEMENTASI ALGORITMA BIRCH DALAM KLASTERISASI KASUS BIAYA HIDUP DI KOTA PADA BEBERAPA NEGARA

Dita Putri Az-Zahra[1], Yusiva Sm Sidabutar [2] , Adhevta Jibran Muzakki Khan[3] , Muhammad Faris[4], Arnita[5]

Universitas negeri Medan

E-mail: azzahradita281@gmail.com[1], yusivasmsidabutar@gmail.com[2] , khanadhevta@gmail.com[3], cherrywax7@gmail.com[4], arnita@unimed.ac.id[5]

Abstrak

Penelitian ini bertujuan untuk mengelompokkan kota-kota di dunia berdasarkan indikator ekonomi seperti biaya hidup, harga sewa, indeks harga restoran, dan daya beli menggunakan metode BIRCH (Balanced Iterative Reducing and Clustering using Hierarchies) dengan bantuan Principal Component Analysis (PCA). Data yang digunakan mencakup berbagai indeks ekonomi yang relevan, termasuk Cost of Living Index, Rent Index, Groceries Index, dan Local Purchasing Power Index. Setelah melakukan preprocessing data, PCA digunakan untuk mengurangi dimensi, dan metode BIRCH diterapkan untuk mengelompokkan kota-kota. Hasilnya, penelitian ini mengidentifikasi tiga klaster utama kota: klaster dengan biaya hidup dan harga sewa rendah, klaster menengah, dan klaster dengan biaya hidup serta harga sewa yang tinggi. Penelitian ini menyimpulkan bahwa perbedaan dalam biaya hidup dan daya beli adalah faktor kunci dalam menentukan pengelompokan kota secara global. Visualisasi dalam bentuk scatter plot dari komponen utama PCA membantu memahami hubungan antar-klaster. Purpose: Tujuan dari penelitian ini adalah untuk mengelompokkan kota-kota di dunia berdasarkan karakteristik ekonomi utama mereka, menggunakan metode BIRCH dan Principal Component Analysis (PCA). Motivasi dari penelitian ini adalah untuk mengidentifikasi dan memahami bagaimana faktor-faktor ekonomi, seperti biaya hidup, harga sewa, dan daya beli, mempengaruhi pengelompokan kota dalam konteks global. Dengan membangun analisis dari dataset yang mengukur biaya hidup dan harga sewa, penelitian ini bertujuan untuk memberikan pemahaman yang lebih jelas tentang bagaimana faktor-faktor ini saling berinteraksi dan menciptakan pengelompokan kota ke dalam beberapa kategori berdasarkan tingkat biaya hidup dan daya beli. Methods/Study design/approach:Penelitian ini menggunakan kombinasi metode Principal Component Analysis (PCA) dan BIRCH Clustering untuk mengklasifikasikan kota-kota berdasarkan faktor ekonomi utama. Data yang digunakan mencakup berbagai indeks ekonomi, seperti Cost of Living Index, Rent Index, Groceries Index, dan Local Purchasing Power Index. Proses pra-pengolahan data melibatkan normalisasi untuk memastikan bahwa setiap indeks dalam skala yang sama. PCA digunakan untuk mengurangi dimensi data menjadi beberapa komponen utama yang merangkum variasi data secara optimal. Setelah itu, metode BIRCH diterapkan untuk melakukan klasterisasi berdasarkan hasil transformasi PCA. Visualisasi scatter plot digunakan untuk membantu memahami hubungan antar-klaster dan memvisualisasikan distribusi kota-kota dalam dua komponen utama PCA. Result/Findings: Penelitian ini berhasil mengidentifikasi tiga klaster utama

[1]*Corresponding author.
Email addresses: azzahradita281@gmail.com , yusivasmsidabutar@gmail.com , khanadhevta@gmail.com , Cherrywax7@gmail.com , arnita@unimed.ac.id
DOI: 10.24114/j-ids.xxxxx

dari kota-kota di dunia berdasarkan karakteristik ekonomi mereka. Hasil klasterisasi menunjukkan bahwa: klaster pertama mencakup kota-kota dengan biaya hidup dan harga sewa yang rendah, klaster kedua mencakup kota-kota dengan biaya hidup sedang dan harga sewa relatif menengah dan klaster ketiga terdiri dari kota-kota dengan biaya hidup dan harga sewa yang tinggi. Dari hasil analisis PCA, ditemukan bahwa biaya hidup dan harga sewa merupakan variabel yang paling berkontribusi dalam membedakan klaster, diikuti oleh indeks restoran dan daya beli lokal. Scatter plot PCA menunjukkan bahwa kota-kota dalam klaster biaya hidup tinggi juga cenderung memiliki harga sewa dan harga barang konsumsi yang tinggi. Hasil ini konsisten dengan teori ekonomi bahwa kota-kota dengan standar hidup yang lebih tinggi memiliki harga sewa dan biaya hidup yang lebih mahal. Novelty/Originality/Value: Keunikan dari penelitian ini terletak pada kombinasi metode Principal Component Analysis (PCA) dan BIRCH Clustering dalam mengelompokkan kota-kota berdasarkan data ekonomi multivariat. Penggunaan PCA sebagai teknik reduksi dimensi memungkinkan interpretasi yang lebih mudah dari variabel-variabel yang saling berhubungan. BIRCH Clustering dipilih karena kemampuannya untuk menangani dataset yang besar dan kompleks, menjadikannya metode yang ideal untuk klasterisasi kota-kota di seluruh dunia berdasarkan biaya hidup. Penelitian ini memberikan kontribusi dengan menyajikan wawasan berbasis data tentang faktor-faktor ekonomi yang membedakan kota-kota di dunia, yang dapat digunakan sebagai referensi untuk analisis ekonomi global dan pengambilan kebijakan yang lebih tepat sasaran.

Kata Kunci — Klasterisasi, Algoritma BIRCH, Data Mining, PCA, Preprocessing Data.

1. PENDAHULUAN

Laju pertumbuhan ekonomi global yang cepat telah mengakibatkan perbedaan biaya hidup di berbagai kota di seluruh dunia. Biaya hidup yang mencakup kebutuhan dasar seperti sewa, harga makanan, serta akses ke layanan kesehatan dan pendidikan, menjadi salah satu aspek penting dalam menentukan kesejahteraan masyarakat di suatu daerah. Perbedaan ini mendorong kebutuhan untuk melakukan analisis lebih dalam mengenai faktor-faktor yang mempengaruhi biaya hidup di berbagai kota, khususnya di beberapa negara yang memiliki karakteristik ekonomi yang beragam. Salah satu pendekatan yang dapat digunakan untuk memahami pola dan variasi biaya hidup di berbagai kota tersebut adalah dengan menggunakan metode klasterisasi [1].

Pendekatan penelitian ini menggunakan metode kuantitatif, karena data yang dianalisis berbasis angka dan statistik. Penelitian ini berfokus pada penerapan teknik klastering untuk mengidentifikasi pola dalam data biaya hidup di kota pada beberapa negara.

Penelitian ini bertujuan untuk mengklasifikasikan data biaya hidup di berbagai kota dari beberapa negara Dengan pendekatan ini, diharapkan dapat ditemukan pola-pola yang relevan dalam distribusi biaya hidup di kota-kota yang berbeda, serta faktor-faktor yang mempengaruhi perbedaan tersebut. Hasil dari penelitian ini diharapkan tidak hanya membantu dalam memahami struktur biaya hidup di berbagai kota, tetapi juga memberikan kontribusi dalam meningkatkan efisiensi pengelolaan ekonomi kota dan perencanaan strategi pembangunan.

Implementasi algoritma BIRCH dalam klasterisasi kasus biaya hidup di berbagai kota dapat memberikan manfaat besar, terutama dalam memahami faktor-faktor utama yang mempengaruhi perbedaan biaya hidup antar kota. Selain itu, hasil klasterisasi dapat digunakan untuk merumuskan kebijakan yang lebih efektif dalam mengelola distribusi ekonomi dan merencanakan pengembangan infrastruktur kota secara lebih tepat. Algoritma ini juga memungkinkan analisis yang lebih cepat dan efisien dalam memproses data biaya hidup yang kompleks, sehingga dapat membantu dalam pengambilan keputusan yang lebih tepat sasaran.

Data mining adalah proses pengolahan data yang bertujuan untuk menggali informasi yang tersimpan dalam dataset. Informasi yang ditemukan melalui penggalian data ini dapat digunakan dalam pengambilan keputusan. Data mining

merupakan sebuah teknik atau bidang ilmu yang relevan dalam berbagai disiplin pengetahuan yang terkait dengan pengolahan data [2].

Clustering adalah adalah metode pada data mining yang digunakan untuk melakukan pengelompokkan data yang memiliki karakteristik mirip ataupun sama menjadi sebuah cluster sedangkan untuk data yang memiliki karakteristik berbeda akan dikelompokkan menjadi cluster baru [3].

Algoritma BIRCH (Balanced Iterative Reducing and Clustering Using Hierarchies) menggunakan konsep Cluster Feature (CF) yang meringkas sebuah informasi mengenai subcluster - subcluster yang ada sehingga dapat mengurangi skala dari sebuah clustering. Algoritma BIRCH memiliki keunggulan dalam menangani data yang besar dan akurat karena dapat mengatasi sebuah outliers atau pencilan. Area kepadatan dari suatu data akan menentukan apakah tersebut dianggap sebagai outlier atau dianggap subcluster. Jika sebuah area padat dengan data, maka akan dianggap sebagai subcluster, sedangkan area yang sangat jarang akan dianggap sebagai outlier. Algoritma BIRCH ini memiliki dua cara kerja yaitu dengan birch scan atau membaca, dimana data yang di bangun akan masuk kesalam sebuah inisial memory CF tree, yang akan dipandang sebagai kompresi multilevel dari data. sedangkan pada cara kerja kedua adalah dengan menggunakan sebuah penyeleksian terhadap algoritma yang akan dipakai dalam bentuk cluster leaf node dari CF tree [4].

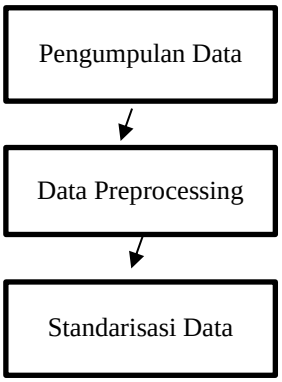
2. METODE

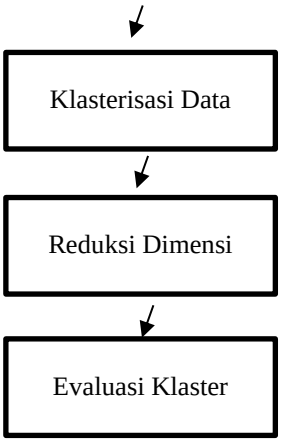
Data Description

Dataset yang digunakan dalam analisis ini merupakan data Cost of Living Index yang diambil dari website Kaggle melalui tautan berikut: Cost of Living Index by Cities | Kaggle. Data ini mengukur biaya hidup di berbagai kota di seluruh dunia, mencakup berbagai indikator yang mempengaruhi kesejahteraan masyarakat. Dataset ini terdiri dari 535 baris data dan 8 kolom, di mana setiap baris mewakili satu kota dengan kolom yang mencatat informasi terkait biaya hidup dan indikator ekonomi lainnya.

Kolom-kolom dalam dataset ini meliputi rank (Peringkat kota berdasarkan biaya hidup, city (nama kota), cost of living Index (Indeks biaya hidup di kota tersebut), rent Index (indeks biaya sewa), Cost of Living Plus Rent Index (Indeks biaya hidup ditambah biaya sewa), Groceries Index (Indeks harga bahan makanan), Restaurant Price Index (indeks harga restoran), dan Local Purchasing Power Index (indeks daya beli lokal). Data ini memberikan wawasan mengenai biaya hidup di masing-masing kota dan faktor-faktor yang mempengaruhinya, yang sangat penting untuk memahami dinamika ekonomi perkotaan.

Jenis variabel yang digunakan dalam dataset ini bervariasi. Kolom seperti City berupa tipe objek (string), sementara kolom lainnya seperti Rent, Restaurant Index, dan Purchasing Power adalah tipe numerik (integer atau float). Dataset ini diharapkan dapat memberikan informasi yang berguna dalam analisis klastering untuk mengelompokkan kota-kota berdasarkan biaya hidup dan faktor-faktor yang memengaruhi kesejahteraan ekonomi warganya.





Data Preprocessing

Pre-processing adalah tahap untuk melakukan transformasi data agar sesuai dengan format seharusnya dan dapat diproses [5].

- Proses pra-pemrosesan data dimulai dengan memeriksa jumlah nilai yang hilang (missing values atau NaN) di setiap kolom. Selanjutnya karena tidak ada missing values atau NaN serta tidak ada duplikat, maka kita tidak perlu melakukan pembersihan data.
- Selanjutnya, dilakukan identifikasi kolom berdasarkan tipe datanya, yakni numerik dan kategorikal. Kolom numerik dapat mencakup atribut seperti 'Biaya Sewa', 'Biaya Makanan', 'Biaya Transportasi', dan 'Pendapatan Rata-rata', sedangkan kolom kategorikal dapat meliputi, 'Kota'. Identifikasi ini penting untuk penerapan teknik encoding yang tepat pada data. Pada data yang digunakan tidak menerapkan encoding dikarenakan hanya satu kolom tipe kategorikal yakni city (kota).
- Outlier atau biasa disebut pencilan merupakan objek pengamatan yang memiliki nilai-nilai ekstrim, baik terlalu rendah atau terlalu tinggi sehingga menyebabkan selisih yang besar dengan objek-objek lainnya. Dalam analisis, outlier biasanya mendapatkan perlakuan khusus supaya tidak menyebabkan masalah dalam penentuan hasil akhir analisis. Tidak semua outlier dapat dihilangkan karena terkadang outlier memiliki nilai yang penting dan berpengaruh dalam kumpulan objek. Terdapat beberapa metode untuk mendeteksi adanya outlier, seperti metode boxplot, grafik, standarisasi, dan jarak kuadrat Mahalanobis.
- Standarisasi data adalah proses untuk menstandarkan data dalam jangkauan yang sama [6]. Data numerik distandardisasi menggunakan StandardScaler dari scikit-learn untuk memastikan bahwa setiap fitur numerik memiliki rata-rata nol dan varians satu, yang penting untuk algoritma klasterisasi yang sensitif terhadap skala data. StandardScaler adalah metode standarisasi fitur dengan menghapus rata-rata (mean) dan menskalakan unit varian (standar deviasi) [7]. Untuk dataset ini, di mana nilai seperti "Cost of Living Index" dan "Rent Index" memiliki skala yang bervariasi tetapi masih dalam kisaran yang relatif sama, standarisasi biasanya lebih cocok, terutama untuk algoritma BIRCH.
- Uji multikolineiritas

Reduksi Dimensi

Tahapan klasterisasi data

4. Klastering

Sebelum melakukan Klastering, dilakukan PCA untuk mengatasi multikolinearitas. Dengan data yang telah diproses dan distandarisasi, klasterisasi dilakukan menggunakan algoritma BIRCH. Algoritma BIRCH merupakan metode pengelompokan yang bersifat hirarki (hierarchical clustering method). Algoritma BIRCH sangat sesuai untuk set data yang sangat besar atau untuk streaming data, karena kemampuannya untuk menemukan solusi pengelompokan yang baik dengan hanya satu pemindaian data [8]. Algoritma

BIRCH dirancang untuk mengelompokkan sejumlah data numerik yang besar dengan mengintegrasikan pengelompokan secara hirarki dan metode pengelompokan lain seperti partisi berulang [9].

5. Alat dan Software

Penelitian ini dilakukan menggunakan perangkat lunak Python dengan bantuan pustaka scikit-learn untuk implementasi algoritma BIRCH dan pustaka pandas untuk manipulasi data. Visualisasi hasil klastering dilakukan menggunakan matplotlib dan seaborn. Python adalah bahasa pemrograman interpretatif multiguna dengan filosofi perancangan yang berfokus pada tingkat keterbacaan kode. Python diklaim sebagai bahasa yang menggabungkan kapabilitas, kemampuan, dengan sintaksis kode yang sangat jelas, dan dilengkapi dengan fungsionalitas pustaka standar yang besar serta komprehensif [10]. Scikit-Learn adalah modul Python yang mengintegrasikan berbagai algoritma pembelajaran mesin untuk masalah yang diawasi dan tidak diawasi skala menengah [11]. Pandas (Python Data Analysis) merupakan pustaka pengolahan data dengan domain pemrograman Python. Pustaka tersebut memiliki fitur data frame yang dapat membantu dalam manajemen data berbasis tabel atau informasi larik (array) [12]. Matplotlib adalah salah satu paket pustaka Python yang paling populer dalam hal visualisasi data. Paket pustaka ini dapat menyajikan model visual plot 2D dari data dalam bentuk array. Pustaka ini juga menyediakan API (Application Programming Interface) yang dapat digunakan dalam aplikasi bantu Python GUI seperti PyQt, WxPython, atau Tkinter. Selain itu juga dapat digunakan dalam shell Python dan IPython, notebook Jupyter, dan web server [13]. Seaborn adalah library yang bertujuan untuk membuat grafik dan statistik dengan menggunakan Python. Seaborn dibangun berdasarkan library matplotlib serta terintegrasi dengan struktur data pada Pandas [14].

3. HASIL DAN PEMBAHASAN

Pada tahap awal, dataset yang digunakan dalam penelitian ini terdiri dari data biaya hidup di kota di beberapa negara yang mencakup atribut seperti kota, biaya hidup, harga sewa, biaya makan dan lainnya. Dengan menggunakan fungsi head(), lima baris pertama dari dataset dapat dilihat pada tabel berikut:

Table 1. baris pertama dataset

Rank	Kota	Cost of Living Index	Rent Index	Groceries Index	
1	Hamilton, Bermuda	137.56	103.03	126.56	
2	Zurich, Switzeland	128.56	62.62	127.35	
3	Basel, Switzerland	126.89	46.14	120.44	
4	Lausanne, Switzerland	119.62	50.35	116.35	
5	Bern, Switzerland	118.42	39.22	114.54	

Dari tabel di atas, dapat dilihat bahwa dataset berisi informasi penting seperti kota, biaya hidup, harga sewa, dan harga grosir.

Table 2. Pengecekan Missing Value

Column	Check for missing value
Rank	0
City	0
Cost of Living Index	0
	0

Rent Index	
Cost of Living Plus Rent Index	0
Groceries Index Restaurant Price Index	0 0
Local Purchaasing Power Index	0

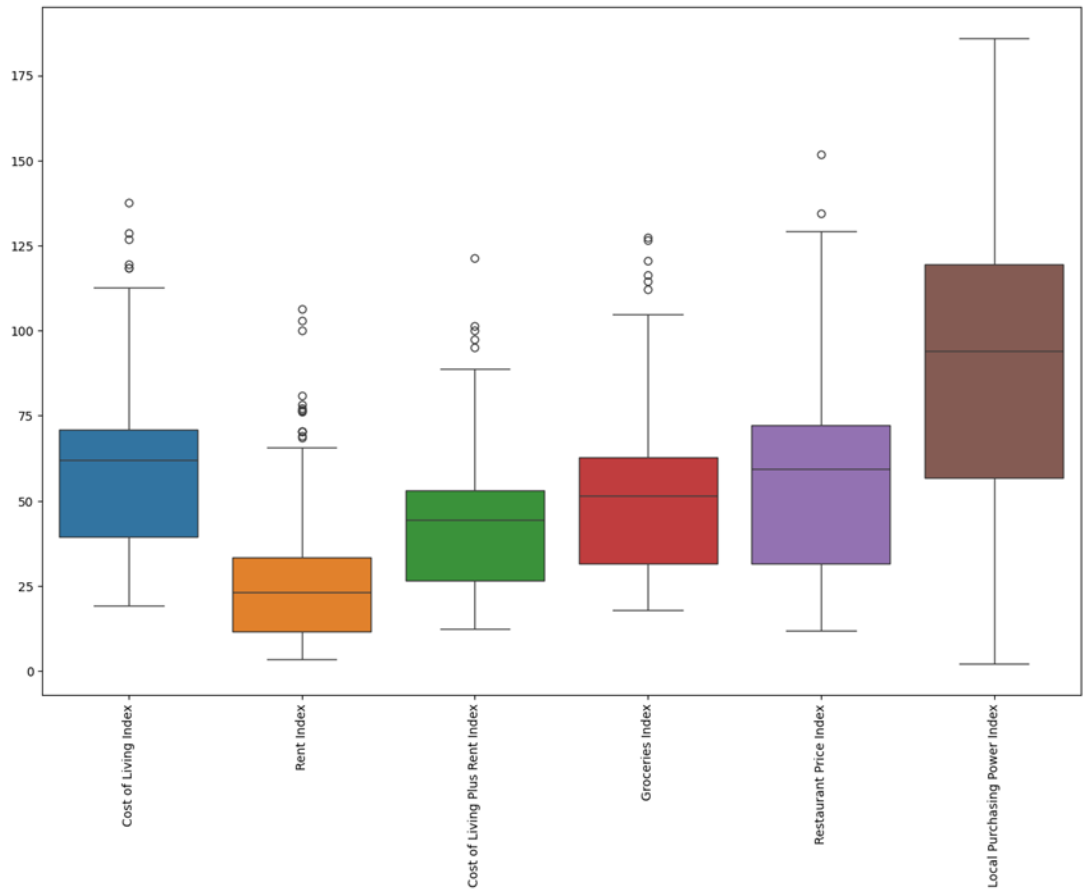
Selain itu, tidak ada nilai yang hilang (missing values) dalam dataset ini, sehingga tidak diperlukan tindakan untuk menangani data yang tidak lengkap.

Data Describe

Table 3. Data Describe

	Rank	Cost of Living Index	Rent Index	Cost of Living Plus Rent Index	Groceries Index	Restaurant Price Index
Count	536.00	536.00	536.00	536.00	536.00	536.00
Mean	268.50	57.187	25.209	42.050	49.758	54.684
Std	154.87	19.984	16.453	17.45	19.320	25.092
Min	1.00	19.260	3.430	12.360	18.01	11.930
25%	134.00	39.370	11.987	26.647	31.497	31.550
50%	268.50	61.970	23.095	44.450	51.570	59.405
75%	402.25	70.865	33.440	53.09	62.757	72.342
max	536.00	137.56	106.49	121.21	127.35	151.77

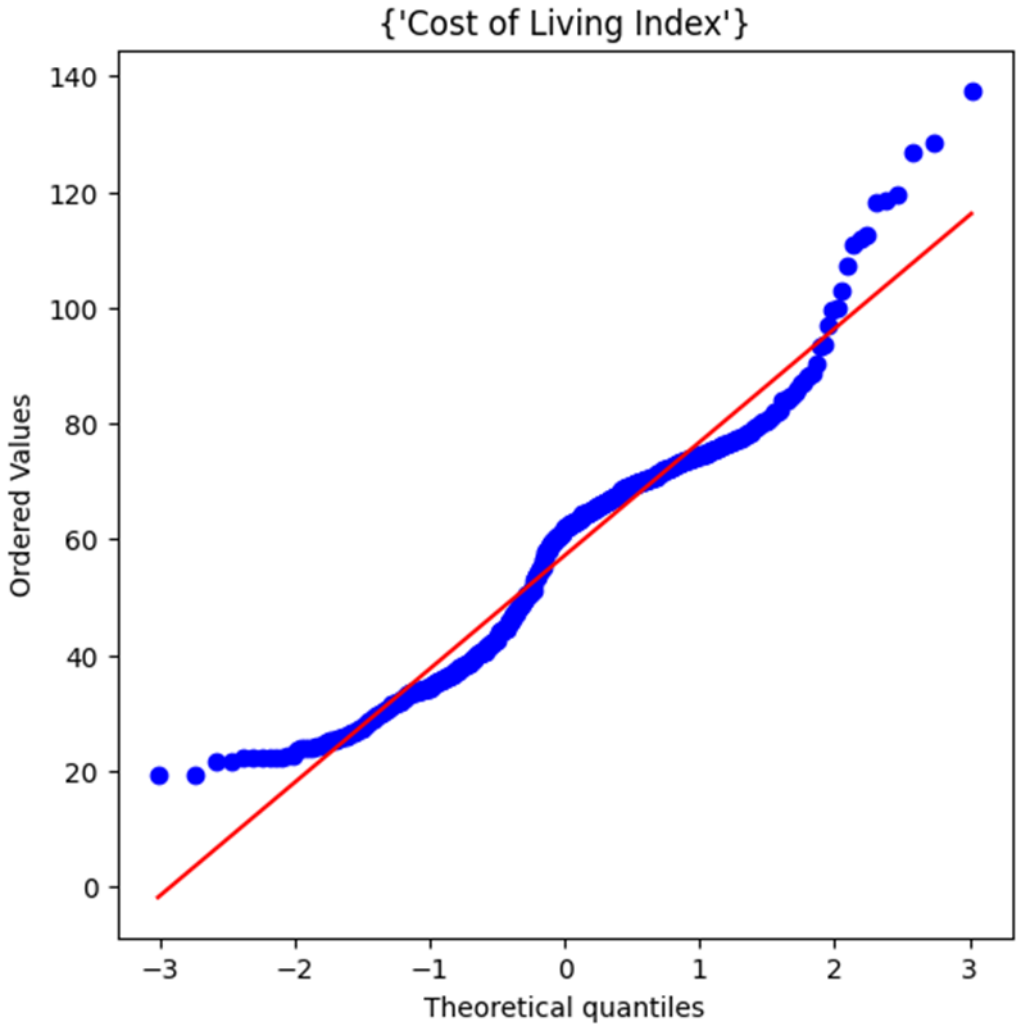
Membuat boxplot untuk setiap fitur



Gambar boxplot yang ditampilkan menggambarkan variasi beberapa indeks terkait biaya hidup dan daya beli di berbagai lokasi. Pada Cost of Living Index, median berada sekitar angka 50, menunjukkan biaya hidup rata-rata di lokasi-lokasi tersebut. Rentangnya cukup lebar di atas median, dengan beberapa outlier yang menandakan adanya tempat dengan biaya hidup yang sangat tinggi. Rent Index menunjukkan variasi yang lebih sempit dibandingkan indeks lainnya, dengan median sekitar 30. Meskipun demikian, ada beberapa lokasi yang memiliki biaya sewa yang lebih tinggi, terlihat dari adanya outlier pada plot tersebut.

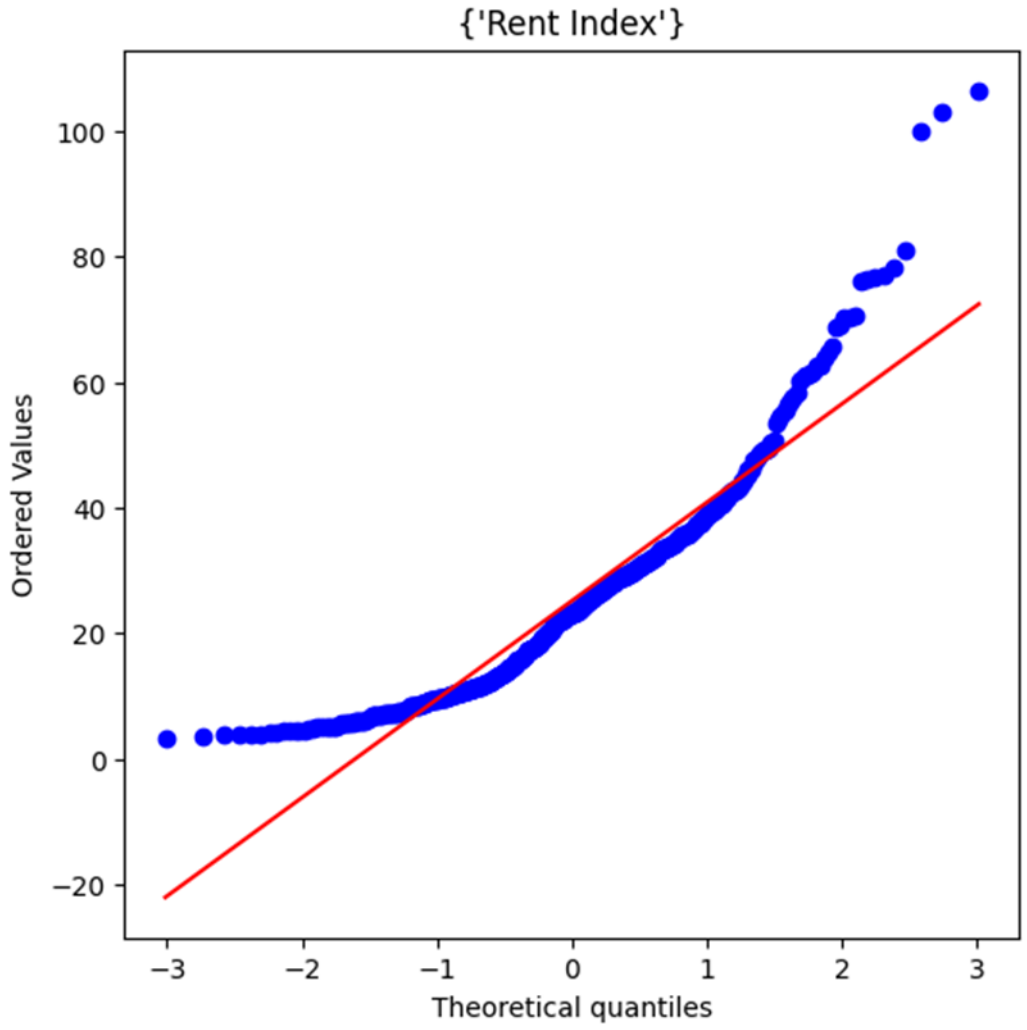
Untuk Cost of Living Plus Rent Index, yang menggabungkan biaya hidup dan sewa, distribusi lebih lebar dibandingkan dengan Rent Index, dengan median mendekati 50. Ini menunjukkan bahwa biaya hidup ditambah sewa di beberapa lokasi lebih bervariasi, terutama pada kuartil atas. Pada Groceries Index, median berada di sekitar 50, yang berarti harga bahan pokok rata-rata tidak jauh berbeda di berbagai lokasi, meskipun ada beberapa tempat dengan harga lebih tinggi yang ditunjukkan oleh adanya outlier.

Restaurant Price Index memiliki median sekitar 60, yang menandakan bahwa makan di restoran cenderung lebih mahal dibandingkan dengan belanja kebutuhan pokok. Namun, terdapat lokasi-lokasi tertentu yang memiliki harga restoran yang jauh lebih mahal, terlihat dari outlier pada boxplot ini. Terakhir, Local Purchasing Power Index menunjukkan variasi yang paling besar dengan median di atas 75. Ini mengindikasikan bahwa daya beli di berbagai lokasi sangat beragam, dengan beberapa tempat memiliki daya beli lokal yang jauh lebih tinggi.



Gambar yang ditampilkan adalah Q-Q plot (Quantile-Quantile plot) untuk fitur **Cost of Living Index**, yang digunakan untuk mengevaluasi apakah data pada fitur tersebut mengikuti distribusi normal. Pada plot ini, titik-titik biru mewakili nilai aktual dari data, sementara garis merah diagonal menunjukkan distribusi normal yang ideal.

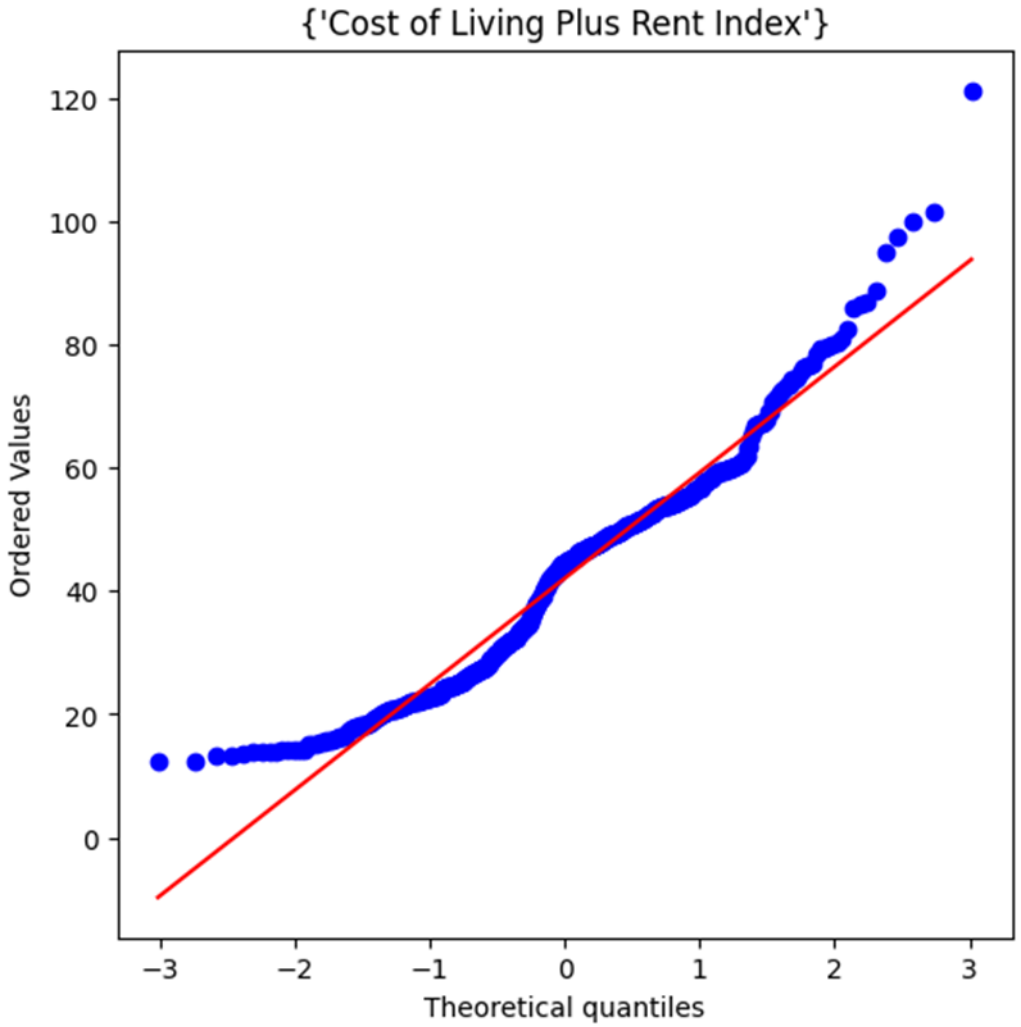
Dari gambar, terlihat bahwa data "Cost of Living Index" cenderung tidak sepenuhnya normal. Di bagian tengah, sebagian besar titik mengikuti garis merah, yang menunjukkan bahwa data di kisaran tersebut cukup mendekati distribusi normal. Namun, di kedua ujungnya, terutama di bagian kanan atas, titik-titik menyimpang cukup jauh dari garis. Ini menunjukkan adanya nilai-nilai ekstrem atau outlier pada data, yang menyebabkan distribusinya tidak sepenuhnya normal. Secara umum, data memiliki beberapa deviasi dari normalitas, terutama di bagian ekor distribusi, di mana outlier menyebabkan pergeseran dari garis teoretis normal.



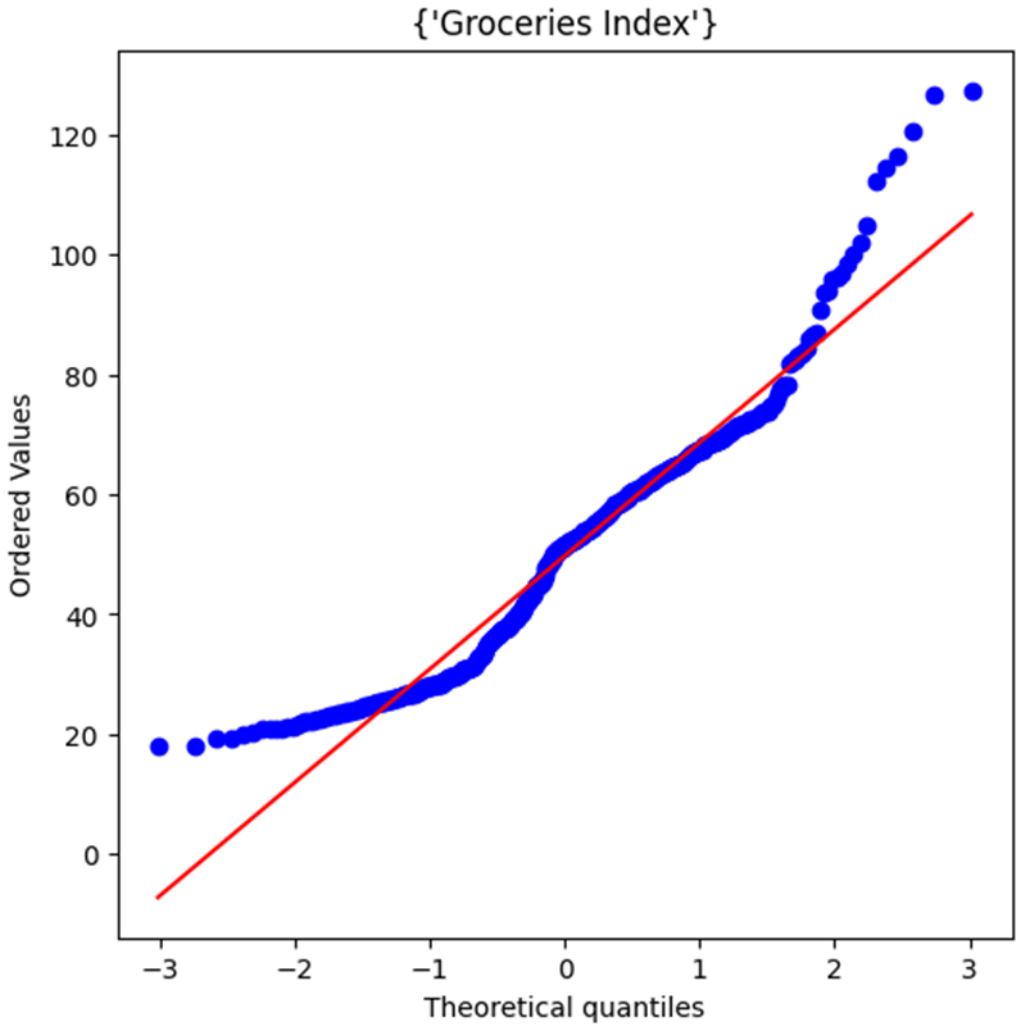
Gambar tersebut adalah Q-Q plot untuk fitur Rent Index, yang digunakan untuk mengevaluasi apakah data pada fitur ini mengikuti distribusi normal. Pada plot ini, titik-titik biru mewakili data aktual, sementara garis merah diagonal menunjukkan distribusi normal ideal.

Dari plot ini, terlihat bahwa data "Rent Index" tidak sepenuhnya mengikuti distribusi normal. Di bagian tengah, sebagian besar titik mendekati garis merah, menunjukkan bahwa nilai-nilai tersebut cukup sesuai dengan distribusi normal. Namun, pada ujung atas, titik-titik mulai menyimpang cukup jauh dari garis, terutama pada nilai kuantile yang tinggi, menunjukkan adanya outlier atau nilai ekstrem pada data sewa di beberapa lokasi.

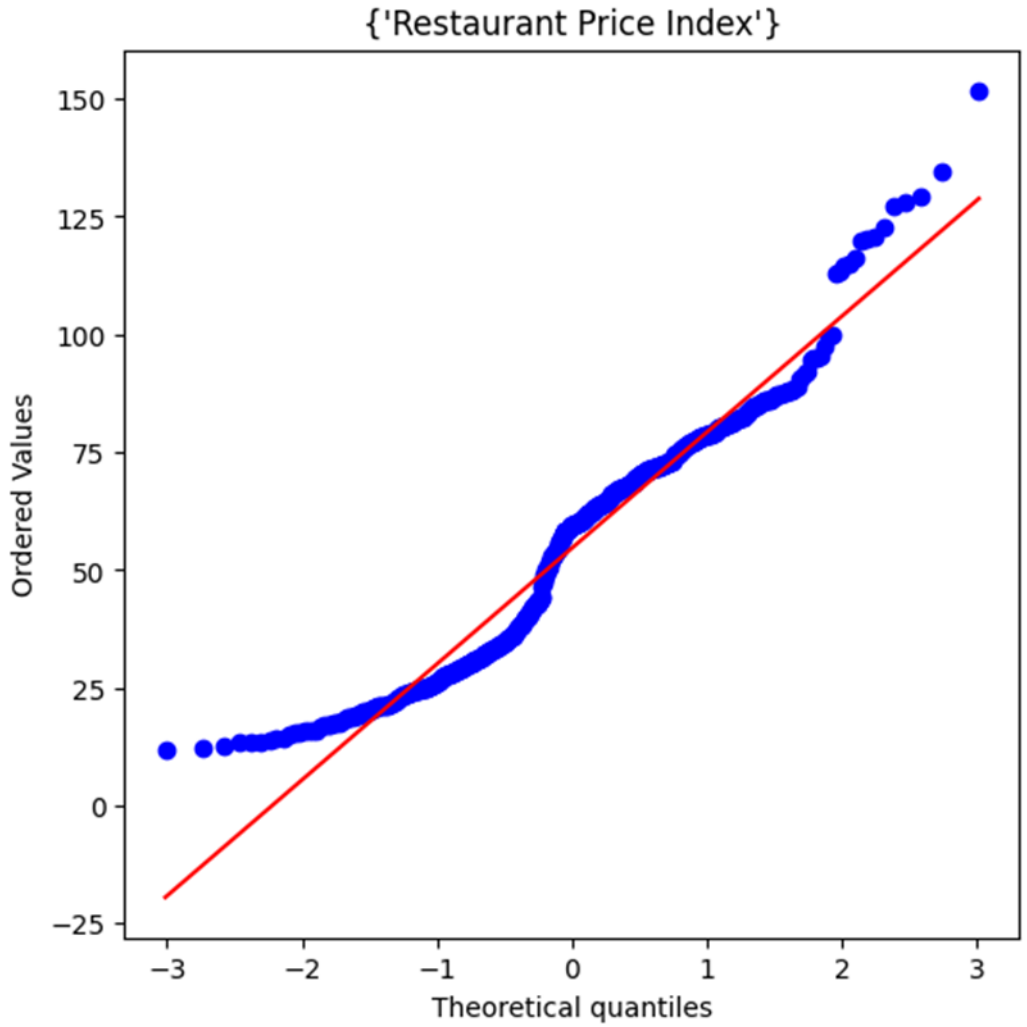
Secara keseluruhan, meskipun sebagian besar data mengikuti pola distribusi normal, terdapat deviasi di bagian ekor kanan plot, menunjukkan bahwa ada beberapa lokasi dengan nilai sewa yang jauh lebih tinggi dari yang diharapkan jika data mengikuti distribusi normal sempurna.



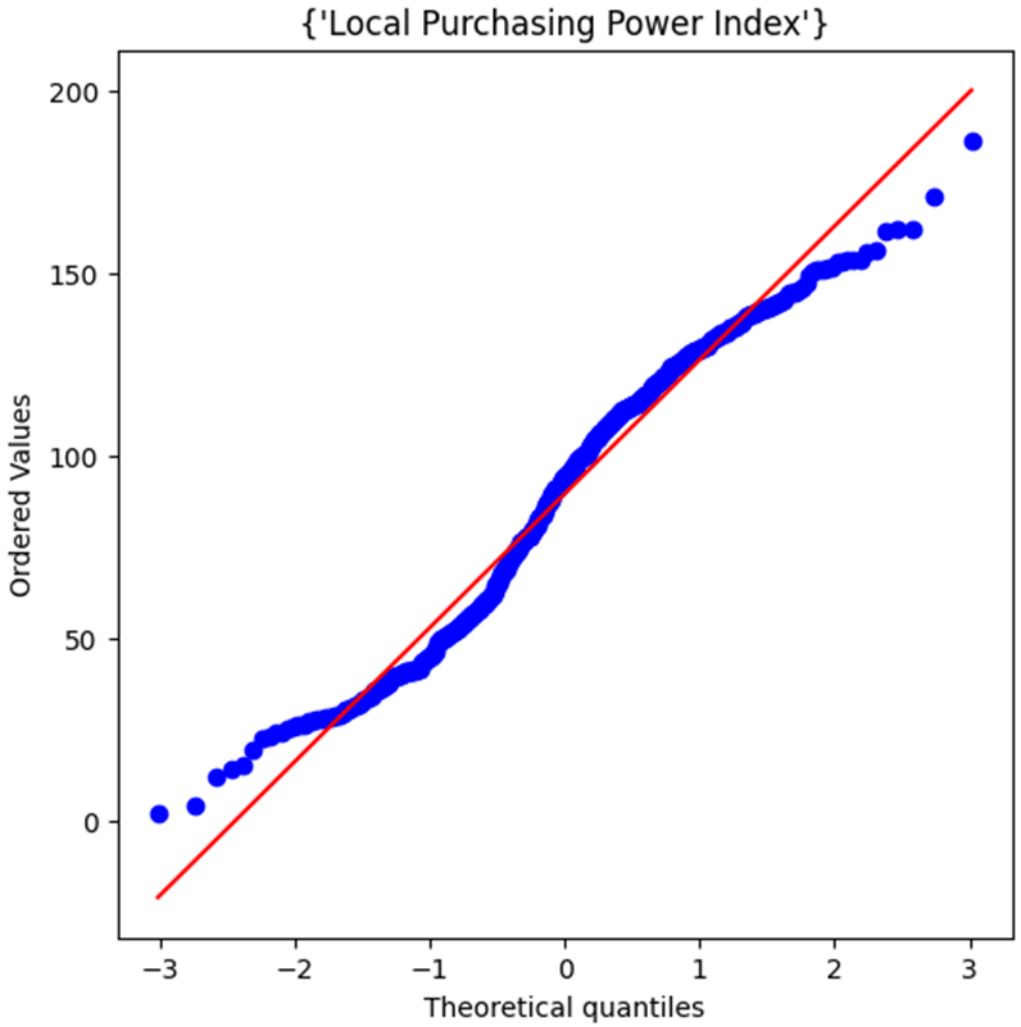
Gambar tersebut merupakan Q-Q plot (Quantile-Quantile plot) yang digunakan untuk membandingkan distribusi data indeks "Cost of Living Plus Rent" dengan distribusi teoretis, biasanya distribusi normal. Pada plot ini, sumbu horizontal (X) mewakili kuantil distribusi teoretis, sedangkan sumbu vertikal (Y) menunjukkan nilai data aktual yang telah diurutkan. Titik-titik biru di grafik menggambarkan perbedaan antara distribusi data dan distribusi teoretis. Jika data mengikuti distribusi normal, titik-titik tersebut akan mengikuti garis merah yang merupakan garis referensi. Namun, pada grafik ini terlihat bahwa pada bagian ekor (nilai rendah dan tinggi), data menyimpang dari garis, menunjukkan adanya deviasi dari distribusi normal, khususnya di bagian nilai yang lebih tinggi (outliers).



Gambar tersebut adalah Q-Q plot yang menggambarkan distribusi data "Groceries Index" dibandingkan dengan distribusi teoretis (normal). Pada grafik ini, sumbu horizontal (X) mewakili kuantil distribusi teoretis, dan sumbu vertikal (Y) menunjukkan nilai urut dari indeks belanja bahan pokok (groceries). Garis merah adalah garis referensi, yang menunjukkan distribusi data jika sepenuhnya mengikuti distribusi normal. Data yang sesuai dengan distribusi normal akan berada di sepanjang garis merah. Namun, pada grafik ini terlihat penyimpangan, terutama di bagian ekor atas (nilai tinggi), yang menunjukkan adanya data ekstrem (outliers) di indeks ini. Ini menandakan bahwa distribusi indeks groceries tidak sepenuhnya normal, terutama karena terdapat beberapa nilai yang jauh lebih tinggi dari yang diharapkan menurut distribusi normal.



Gambar tersebut adalah Q-Q plot yang menunjukkan perbandingan antara distribusi data "Restaurant Price Index" dengan distribusi teoretis (normal). Sumbu horizontal (X) menggambarkan kuantil dari distribusi normal teoretis, sementara sumbu vertikal (Y) menampilkan nilai urut dari indeks harga restoran. Garis merah merupakan garis referensi yang menunjukkan di mana data akan berada jika distribusi mengikuti distribusi normal. Dalam hal ini, titik-titik biru sebagian besar berada di sepanjang garis referensi, tetapi terlihat penyimpangan pada bagian ekor atas, di mana terdapat beberapa nilai yang lebih tinggi dari yang diharapkan (outliers). Hal ini menunjukkan bahwa distribusi indeks harga restoran memiliki deviasi dari distribusi normal, dengan beberapa kota memiliki harga restoran yang sangat tinggi.

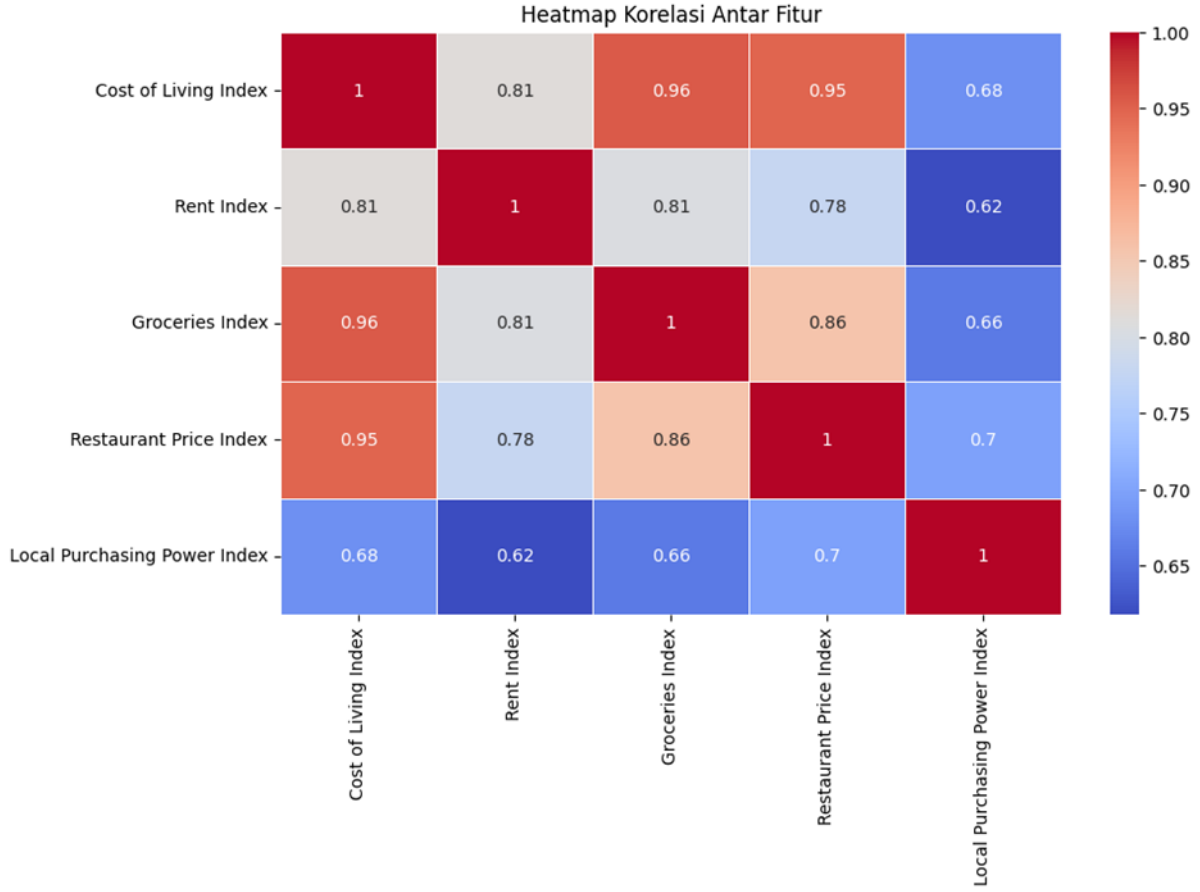


Gambar di atas menunjukkan QQ plot (Quantile-Quantile plot) dari Local Purchasing Power Index. QQ plot adalah alat statistik yang digunakan untuk membandingkan distribusi dari suatu data dengan distribusi teoritis (biasanya distribusi normal). Pada sumbu horizontal (teoretis kuantil) ditampilkan nilai-nilai kuantil dari distribusi normal, sedangkan sumbu vertikal (ordered values) menunjukkan nilai-nilai yang terurut dari data aktual Local Purchasing Power Index.

Garis merah pada plot ini menggambarkan distribusi normal teoretis. Jika data mengikuti distribusi normal, maka titik-titik pada plot seharusnya mengikuti garis merah ini dengan baik. Dari gambar, terlihat bahwa data Local Purchasing Power Index sebagian besar mengikuti garis teoretis ini, meskipun terdapat penyimpangan pada ekor (baik di bagian bawah maupun atas) yang menunjukkan adanya beberapa nilai outlier atau data yang mungkin lebih menyimpang dari distribusi normal pada ekstrem-ekstremnya.

Secara keseluruhan, QQ plot ini mengindikasikan bahwa Local Purchasing Power Index memiliki distribusi yang mendekati normal, tetapi terdapat beberapa anomali yang mungkin perlu diperhatikan lebih lanjut dalam analisis mendalam.

Heatmap Korelasi Antar Fitur

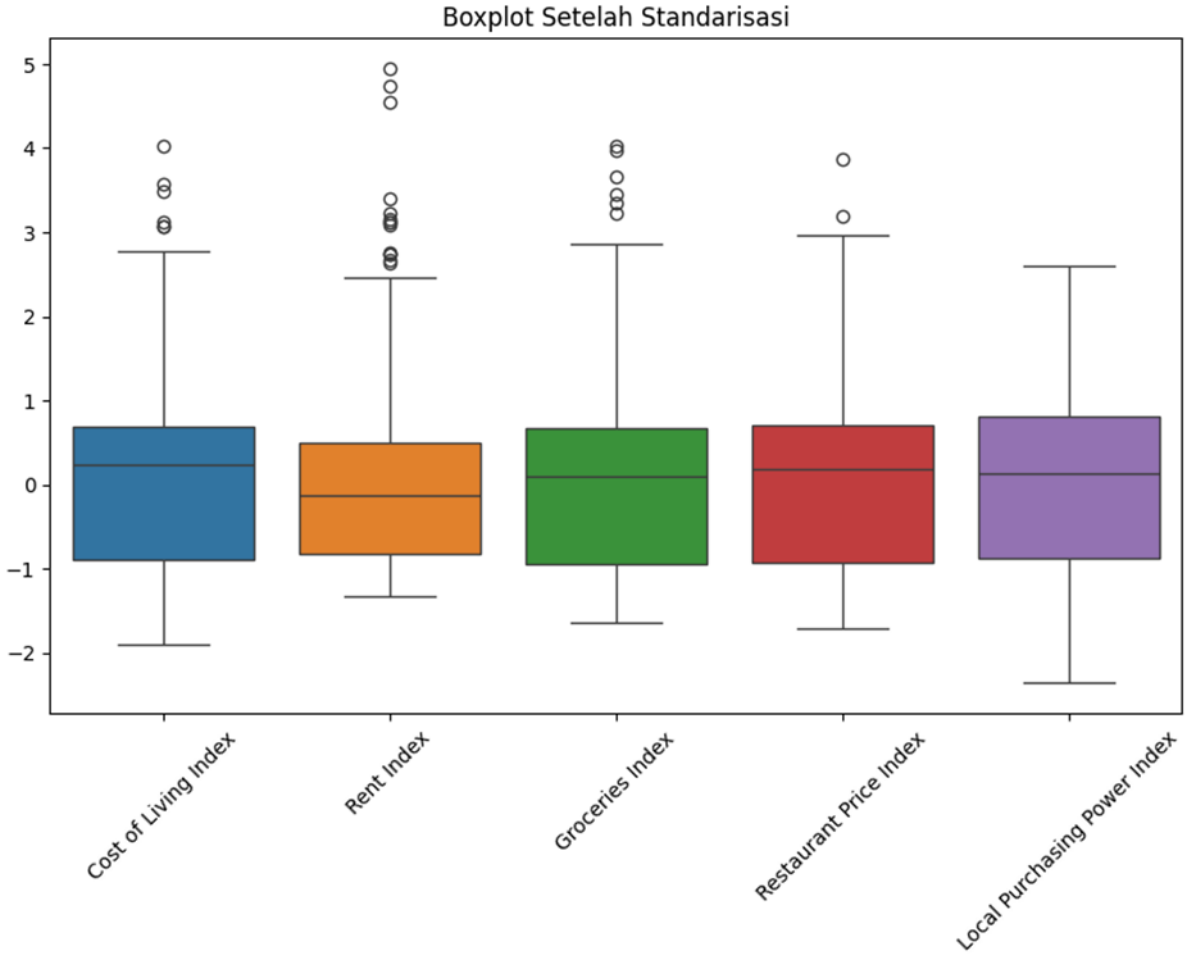


Gambar di atas menunjukkan heatmap korelasi antar fitur yang mengilustrasikan hubungan antara beberapa indeks ekonomi, yaitu Cost of Living Index, Rent Index, Groceries Index, Restaurant Price Index, dan Local Purchasing Power Index. Heatmap ini menggunakan koefisien korelasi Pearson, yang menunjukkan seberapa kuat dan searah hubungan antar variabel. Nilai korelasi berkisar dari -1 hingga 1, di mana nilai 1 menunjukkan korelasi sempurna positif, -1 menunjukkan korelasi sempurna negatif, dan 0 berarti tidak ada korelasi.

Dari heatmap ini, dapat disimpulkan beberapa poin penting:

1. Cost of Living Index memiliki korelasi sangat kuat dengan Groceries Index (0,96) dan Restaurant Price Index (0,95). Ini menunjukkan bahwa biaya hidup sangat dipengaruhi oleh harga kebutuhan sehari-hari seperti bahan makanan dan makan di restoran.
2. Rent Index memiliki korelasi sedang dengan Cost of Living Index (0,81), yang menandakan bahwa sewa tempat tinggal merupakan salah satu komponen penting dalam biaya hidup, meskipun tidak sekuat pengaruh harga makanan.
3. Local Purchasing Power Index memiliki korelasi positif yang relatif lebih rendah dengan semua indeks lainnya, terutama dengan Cost of Living Index (0,68), yang menunjukkan bahwa daya beli lokal tidak secara langsung berbanding lurus dengan biaya hidup.

Boxplot Setelah Standarisasi

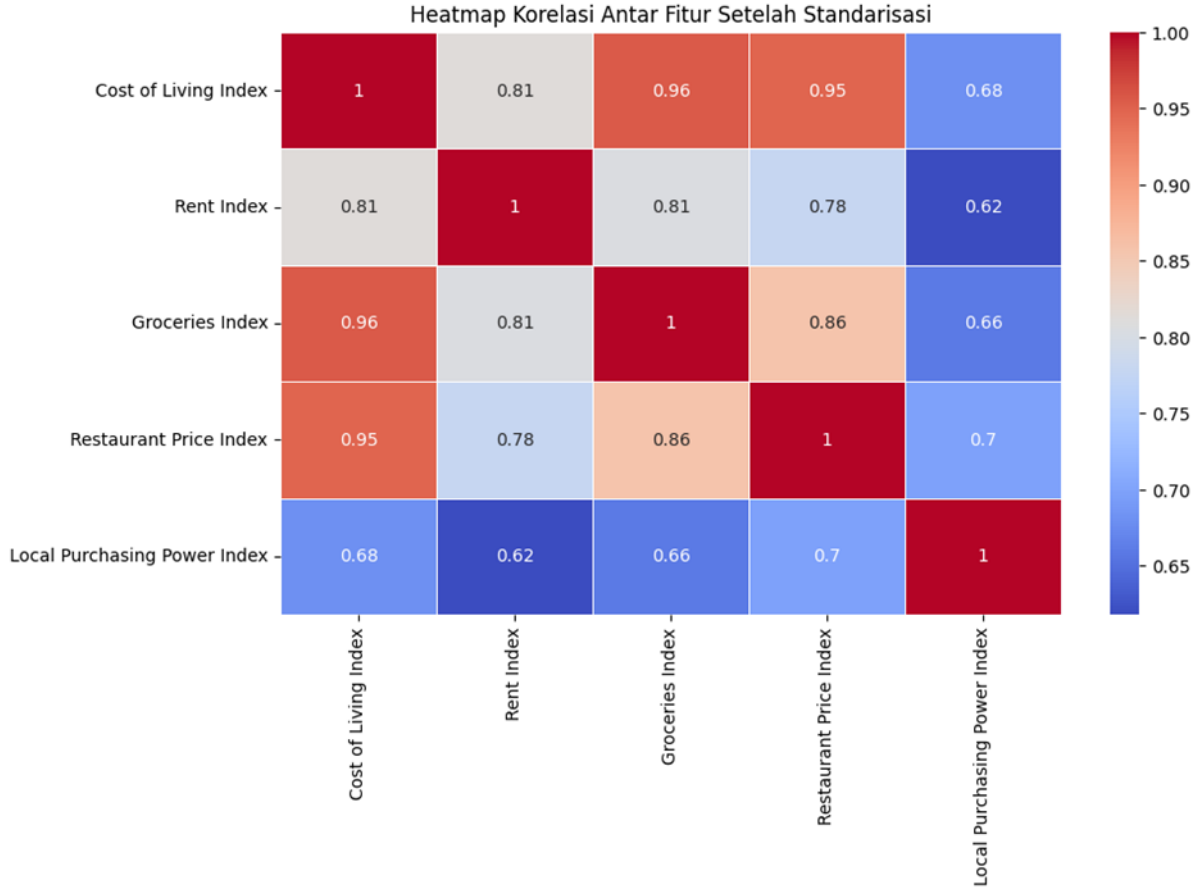


Gambar di atas menunjukkan boxplot dari beberapa indeks ekonomi setelah dilakukan proses standarisasi. Boxplot digunakan untuk memberikan gambaran distribusi data, seperti rentang, kuartil, dan keberadaan outlier.

Beberapa poin penting yang dapat diambil dari boxplot ini adalah sebagai berikut:

1. Cost of Living Index: Distribusi data untuk indeks ini cukup simetris dengan sedikit outlier yang terdeteksi di bagian atas. Median berada di atas nol, menunjukkan bahwa mayoritas nilai indeks ini lebih tinggi dari rata-rata.
2. Rent Index: Memiliki rentang yang lebih sempit dibandingkan dengan indeks lainnya, tetapi terdapat banyak outlier di bagian atas. Hal ini menunjukkan bahwa beberapa lokasi memiliki nilai sewa yang jauh lebih tinggi dibandingkan rata-rata.
3. Groceries Index: Distribusi data menunjukkan median yang mendekati nol dengan beberapa outlier di bagian atas. Rentang dari kuartil pertama hingga kuartil ketiga lebih besar daripada Rent Index, yang menunjukkan variasi yang lebih besar dalam harga bahan makanan.
4. Restaurant Price Index: Memiliki pola distribusi yang mirip dengan Groceries Index, di mana terdapat beberapa outlier di bagian atas. Median sedikit lebih tinggi dari nol, menandakan bahwa harga di restoran umumnya lebih tinggi daripada rata-rata yang telah distandarisasi.
5. Local Purchasing Power Index: Memiliki distribusi yang relatif seimbang, dengan median berada di atas nol dan tidak ada outlier yang signifikan. Rentang variabilitas indeks ini lebih kecil daripada indeks lainnya, yang menunjukkan bahwa daya beli lokal cenderung lebih konsisten dibandingkan harga-harga barang dan jasa lainnya.

Heatmap Korelasi Antar Fitur Setelah Standarisasi



Gambar di atas menunjukkan heatmap korelasi antar fitur setelah standarisasi yang menggambarkan hubungan antar variabel indeks ekonomi, yaitu Cost of Living Index, Rent Index, Groceries Index, Restaurant Price Index, dan Local Purchasing Power Index. Korelasi diukur menggunakan koefisien Pearson, dengan skala dari -1 hingga 1. Nilai positif mendekati 1 menunjukkan korelasi kuat dan searah, sedangkan nilai mendekati -1 menunjukkan korelasi kuat namun berlawanan arah.

Beberapa poin penting dari heatmap ini adalah sebagai berikut:

1. Cost of Living Index memiliki korelasi sangat kuat dengan Groceries Index (0,96) dan Restaurant Price Index (0,95). Hal ini mengindikasikan bahwa semakin tinggi biaya hidup, harga bahan makanan dan harga di restoran juga cenderung meningkat secara signifikan.
2. Rent Index memiliki korelasi moderat dengan Cost of Living Index (0,81) dan Groceries Index (0,81), yang menandakan bahwa sewa tempat tinggal memberikan kontribusi penting pada biaya hidup, tetapi tidak sekuat pengaruh harga makanan.
3. Local Purchasing Power Index memiliki korelasi yang lebih rendah dengan fitur lainnya, dengan korelasi tertinggi dengan Cost of Living Index (0,68) dan Restaurant Price Index (0,7). Ini menunjukkan bahwa daya beli lokal cenderung tidak terlalu erat berhubungan dengan biaya hidup atau harga-harga kebutuhan pokok, tetapi tetap ada hubungan moderat antara daya beli dengan harga restoran dan biaya hidup secara keseluruhan.

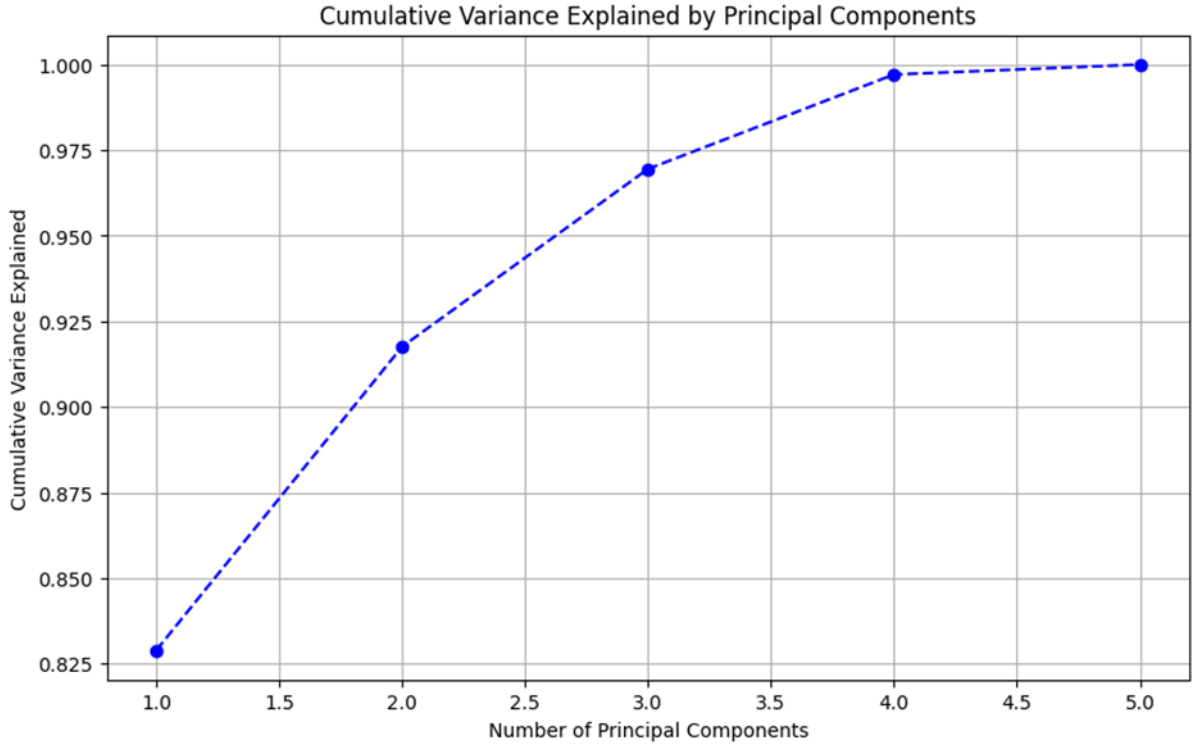
Heatmap ini memberikan wawasan bahwa meskipun ada hubungan yang signifikan antara beberapa indeks ekonomi, daya beli lokal tampaknya lebih mandiri dan tidak sepenuhnya dipengaruhi oleh faktor-faktor seperti sewa atau harga bahan makanan. Standarisasi membantu memperjelas hubungan antar fitur dengan menjaga semua variabel pada skala yang sama.

Cumulative Variance Explained by Principal Components

```
Variansi yang dijelaskan oleh setiap komponen utama (PC):
PC1: 82.89%
PC2: 8.87%
PC3: 5.19%
PC4: 2.76%
PC5: 0.29%

Data setelah PCA:
        PC1       PC2       PC3       PC4       PC5                  City
0  7.886860  2.648374  0.765711 -0.282271 -0.081423     Hamilton, Bermuda
1  6.465422  1.201740 -0.835352  0.772658 -0.061490    Zurich, Switzerland
2  5.848489  0.894269 -1.631131  0.580075 -0.014637     Basel, Switzerland
3  5.477423  1.023283 -1.185891  0.542792 -0.112128  Lausanne, Switzerland
4  4.783728  1.223512 -1.651223  0.736323 -0.011326      Bern, Switzerland
```

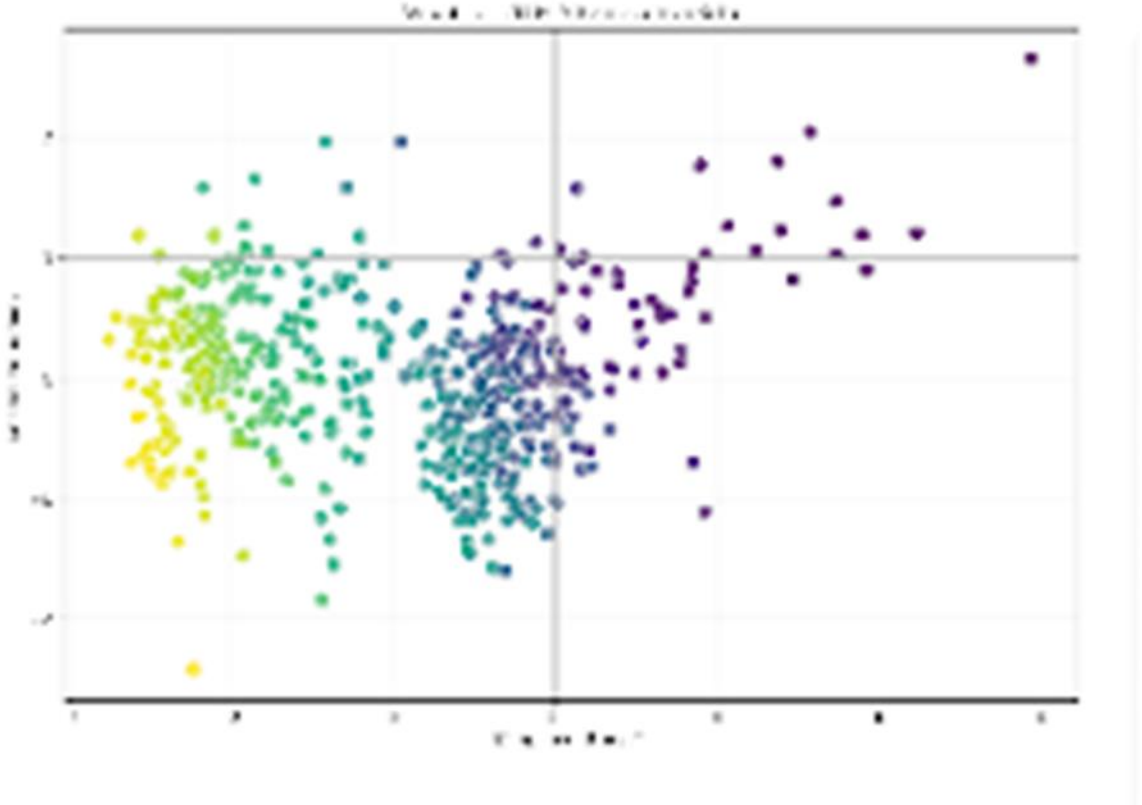
Gambar tersebut menunjukkan variansi yang dijelaskan oleh setiap komponen utama (Principal Components atau PC) dalam analisis PCA (Principal Component Analysis). PC1 menjelaskan sebagian besar variansi dalam data, yaitu sebesar 82,89%. PC2 menjelaskan variansi tambahan sebesar 8,87%, diikuti oleh PC3 dengan 5,19%, PC4 dengan 2,76%, dan PC5 hanya menyumbang 0,29%. Secara kumulatif, PC1 hingga PC5 memberikan gambaran yang cukup lengkap dari variansi total dalam dataset. Setelah PCA dilakukan, data diproyeksikan ke dalam komponen utama (PC1 hingga PC5), dengan nilai-nilai baru yang menunjukkan representasi data di ruang dimensi yang lebih rendah. Setiap baris dalam data setelah PCA menggambarkan nilai dari masing-masing PC untuk kota tertentu, seperti Hamilton, Zurich, Basel, Lausanne, dan Bern.



Gambar tersebut menunjukkan grafik Cumulative Variance Explained oleh Principal Components (PC) yang menggambarkan akumulasi variansi yang dijelaskan oleh setiap komponen utama. Sumbu horizontal mewakili jumlah komponen utama, sementara

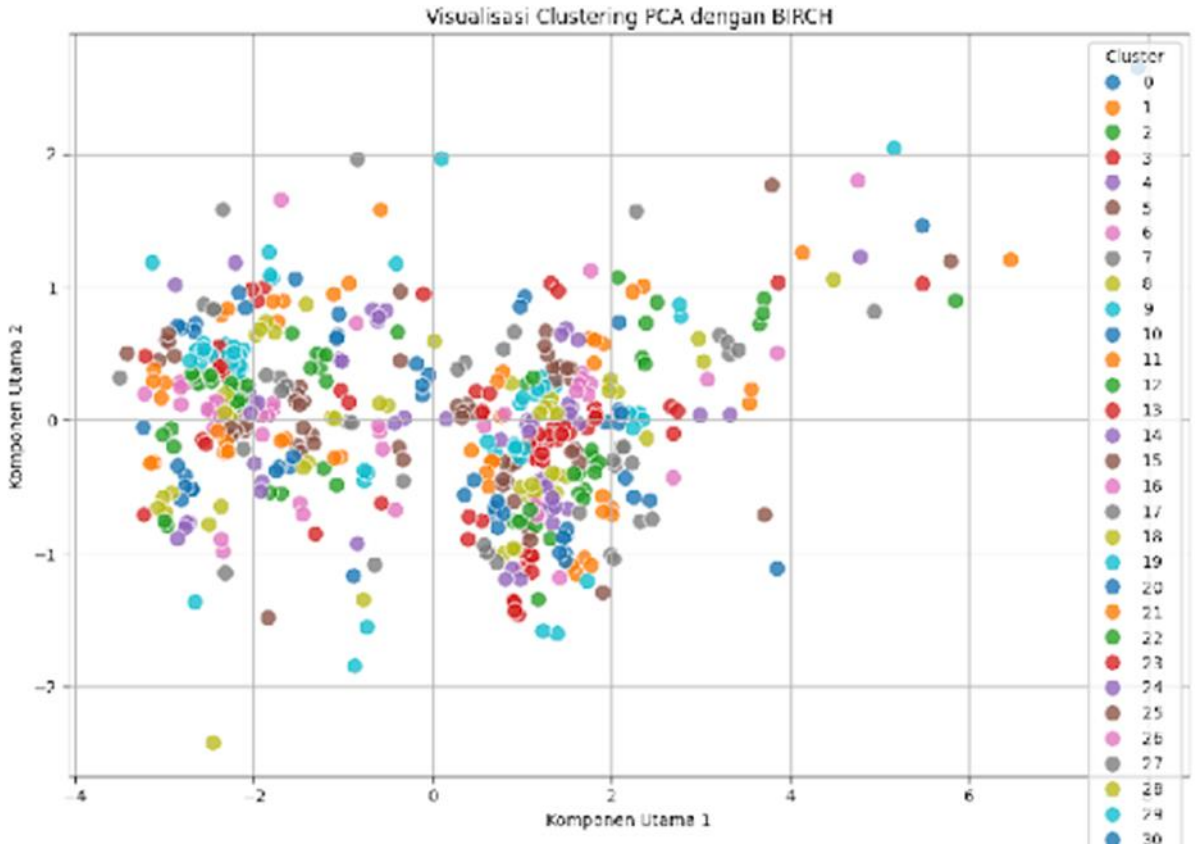
sumbu vertikal menunjukkan persentase kumulatif dari total variansi yang dijelaskan. Dari grafik, terlihat bahwa PC1 menjelaskan sekitar 82,5% variansi. Ketika ditambahkan PC2, kumulatif variansi yang dijelaskan meningkat hingga mendekati 91%. Penambahan PC3, PC4, dan PC5 secara bertahap meningkatkan variansi kumulatif yang dijelaskan hingga mencapai hampir 100%. Ini menunjukkan bahwa sebagian besar informasi dalam dataset dapat dijelaskan dengan hanya menggunakan beberapa komponen utama pertama, terutama PC1 dan PC2, sehingga PCA membantu dalam mengurangi dimensi data tanpa kehilangan banyak informasi.

Visualisasi 2D PCA Berdasarkan Kota



Gambar tersebut menunjukkan visualisasi dua dimensi dari hasil analisis PCA (Principal Component Analysis) berdasarkan kota. Visualisasi ini menggunakan dua komponen utama (Komponen Utama 1 dan Komponen Utama 2) untuk merepresentasikan data dalam ruang 2D. Setiap titik dalam grafik mewakili satu kota, dengan warna yang berbeda menunjukkan variasi antara kota-kota tersebut. Komponen Utama 1 (di sumbu horizontal) dan Komponen Utama 2 (di sumbu vertikal) memungkinkan kita melihat pola pengelompokan di antara kota-kota yang mirip. Dari gambar, terlihat bahwa kota-kota dengan karakteristik yang serupa cenderung berkelompok bersama, yang mengindikasikan bahwa PCA berhasil mereduksi dimensi data dengan tetap mempertahankan pola utama dalam data.

Visualisasi Clustering PCA dengan BIRCH



Gambar di atas merupakan visualisasi hasil clustering menggunakan algoritma Balanced Iterative Reducing and Clustering using Hierarchies (BIRCH) yang dioptimalkan dengan Principal Component Analysis (PCA). Visualisasi ini menampilkan data yang telah direduksi menjadi dua komponen utama melalui PCA, yang ditampilkan pada sumbu X dan Y sebagai "Komponen Utama 1" dan "Komponen Utama 2". Setiap titik pada grafik mewakili satu data, dan warna titik menunjukkan kluster yang berbeda hasil dari proses clustering.

Legenda di sebelah kanan menggambarkan jumlah kluster yang dihasilkan oleh BIRCH, dengan masing-masing kluster diberi warna unik dari 0 hingga 30. Tujuan dari PCA adalah untuk mereduksi dimensi data sehingga lebih mudah divisualisasikan, sementara algoritma BIRCH digunakan untuk mengelompokkan data berdasarkan kemiripannya.

Secara umum, titik-titik yang memiliki warna yang sama berada berdekatan satu sama lain, yang menunjukkan bahwa data dalam kluster tersebut memiliki karakteristik yang serupa. Visualisasi ini membantu dalam memahami distribusi data setelah proses clustering, serta hubungan antar kluster yang mungkin terlihat lebih jelas di ruang dimensi yang lebih rendah.

Profil Klaster

Profil Klaster:

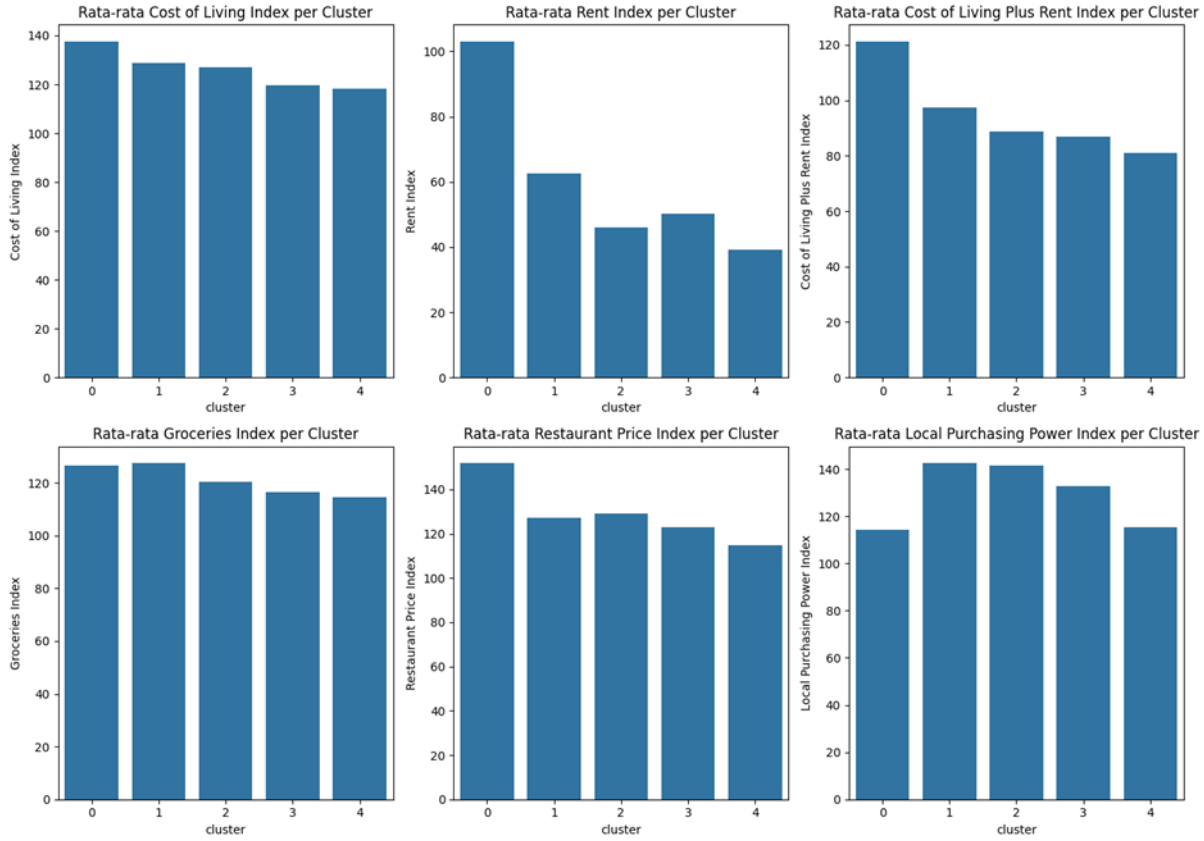
	cluster	Cost of Living Index	Rent Index	Cost of Living Plus Rent Index	\
0	0	137.56	103.03	121.21	
1	1	128.65	62.62	97.39	
2	2	126.89	46.14	88.66	
3	3	119.62	50.35	86.83	
4	4	118.42	39.22	80.93	

	Groceries Index	Restaurant Price Index	Local Purchasing Power Index
0	126.56	151.77	114.19
1	127.35	127.14	142.39
2	120.44	129.10	141.48
3	116.35	122.83	132.58
4	114.54	114.86	115.48

Gambar tersebut menunjukkan ringkasan analisis klaster berdasarkan berbagai indeks yang terkait dengan biaya hidup. Data dibagi menjadi lima klaster (0 hingga 4). Setiap klaster memiliki karakteristik berbeda berdasarkan nilai beberapa indeks berikut:

1. **Cost of Living Index (Indeks Biaya Hidup):** Indeks ini mengukur biaya hidup umum di setiap klaster. Klaster 0 memiliki biaya hidup tertinggi sebesar 137.56, sedangkan Klaster 4 memiliki biaya hidup terendah sebesar 118.42.
2. **Rent Index (Indeks Sew**a): Indeks ini menunjukkan rata-rata biaya sewa. Klaster 0 juga memiliki indeks sewa tertinggi (103.03), menunjukkan bahwa biaya sewa merupakan faktor signifikan dalam biaya hidup di klaster ini. Di sisi lain, Klaster 4 memiliki indeks sewa terendah (39.22), yang menunjukkan biaya sewa yang lebih rendah.
3. **Cost of Living Plus Rent Index (Indeks Biaya Hidup Plus Sewa):** Indeks ini menggabungkan biaya hidup dan sewa menjadi satu. Klaster 0 memiliki nilai tertinggi yaitu 121.21, menunjukkan bahwa total pengeluaran di klaster ini adalah yang tertinggi, sedangkan Klaster 4 memiliki nilai terendah yaitu 80.93.
4. **Groceries Index (Indeks Bahan Makanan):** Indeks ini mencerminkan biaya bahan makanan. Klaster 0 memiliki indeks bahan makanan tertinggi (126.56), sedangkan Klaster 4 memiliki yang terendah (114.54), menunjukkan adanya variasi harga bahan makanan di berbagai klaster.
5. **Restaurant Price Index (Indeks Harga Restoran):** Indeks ini mengukur biaya makan di luar. Klaster 0 kembali memiliki nilai tertinggi yaitu 151.77, yang menunjukkan bahwa makan di luar cukup mahal di sana. Klaster 4 memiliki indeks harga restoran terendah (114.86).
6. **Local Purchasing Power Index (Indeks Daya Beli Lokal):** Indeks ini menunjukkan daya beli penduduk lokal. Klaster 1 memiliki daya beli tertinggi (142.39), yang berarti bahwa orang-orang di klaster ini dapat membeli lebih banyak barang dan jasa relatif terhadap pendapatan mereka. Sebaliknya, Klaster 4 memiliki daya beli yang relatif lebih rendah (115.48).

Rata-rata Local Purchasing Power Index per Cluster



Berikut adalah penjelasan lengkap dari semua grafik yang ditampilkan:

1. Rata-rata Cost of Living Index per Cluster

Grafik ini menunjukkan rata-rata indeks biaya hidup di setiap klaster. Klaster 0 memiliki biaya hidup tertinggi dengan indeks mendekati 140, menunjukkan bahwa klaster ini memiliki harga barang dan jasa yang lebih tinggi secara umum. Sementara itu, Klaster 4 memiliki indeks biaya hidup terendah, berada di sekitar 118, yang menunjukkan bahwa klaster ini memiliki biaya hidup yang lebih rendah dibandingkan klaster lainnya.

2. Rata-rata Rent Index per Cluster

Grafik ini menggambarkan rata-rata indeks sewa di setiap klaster. Klaster 0 memiliki indeks sewa tertinggi, lebih dari 100, menunjukkan bahwa biaya sewa di klaster ini sangat tinggi. Sebaliknya, Klaster 4 memiliki indeks sewa terendah, kurang dari 40, yang berarti biaya sewa di klaster ini jauh lebih terjangkau.

3. Rata-rata Cost of Living Plus Rent Index per Cluster

Grafik ini menggabungkan biaya hidup dan sewa menjadi satu indeks. Klaster 0 menonjol dengan indeks tertinggi, mendekati 120, yang menunjukkan bahwa total biaya hidup, termasuk sewa, adalah yang tertinggi di klaster ini. Klaster 4 memiliki indeks terendah, sekitar 80, menunjukkan bahwa total pengeluaran untuk biaya hidup dan sewa adalah yang paling rendah di klaster ini.

4. Rata-rata Groceries Index per Cluster

Grafik ini menunjukkan rata-rata indeks bahan makanan di setiap klaster. Klaster 0 memiliki indeks bahan makanan tertinggi, lebih dari 120, yang menunjukkan bahwa harga bahan makanan di klaster ini relatif tinggi. Klaster 4 memiliki indeks terendah, sedikit di atas 110, menunjukkan bahwa bahan makanan di klaster ini relatif lebih terjangkau dibandingkan dengan klaster lainnya.

5. Rata-rata Restaurant Price Index per Cluster

Grafik ini menggambarkan rata-rata indeks harga restoran di setiap klaster. Klaster 0 kembali memiliki indeks tertinggi, mendekati 150, yang menunjukkan bahwa biaya makan di luar di klaster ini cukup mahal. Sebaliknya, Klaster 4 memiliki indeks restoran terendah, sekitar 115, menunjukkan bahwa biaya makan di restoran lebih murah di klaster ini.

6. Rata-rata Local Purchasing Power Index per Cluster

Grafik ini menampilkan daya beli lokal rata-rata di setiap klaster. Klaster 1 memiliki daya beli tertinggi dengan indeks mendekati 145, yang berarti penduduk di klaster ini memiliki kemampuan untuk membeli lebih banyak barang dan jasa relatif terhadap pendapatan mereka. Klaster 4 memiliki daya beli lokal terendah, sekitar 115, menunjukkan bahwa penduduk di klaster ini memiliki kemampuan membeli yang lebih rendah, mungkin karena pendapatan yang lebih rendah atau biaya hidup yang lebih tinggi.

```
Silhouette Score: 0.33
Calinski-Harabasz Index: 1570.32
Davies-Bouldin Index: 0.46
```

Gambar tersebut menunjukkan tiga metrik evaluasi yang digunakan untuk menilai kualitas hasil klasterisasi, yaitu Silhouette Score, Calinski-Harabasz Index, dan Davies-Bouldin Index:

1. Silhouette Score: Nilai ini adalah 0,33. Silhouette Score berkisar antara -1 dan 1, di mana nilai yang lebih mendekati 1 menunjukkan bahwa klaster-klaster tersebut terbentuk dengan baik dan terpisah dengan jelas. Nilai 0,33 menunjukkan bahwa klaster-klaster tersebut tidak memiliki pemisahan yang sangat jelas, meskipun masih ada beberapa tingkat pemisahan antara klaster.
2. Calinski-Harabasz Index: Nilainya adalah 1570,32. Indeks ini mengukur rasio antara jumlah dispersi di dalam klaster dengan dispersi antar klaster. Semakin tinggi nilai indeks ini, semakin baik kualitas klasterisasi. Dengan nilai yang cukup tinggi seperti 1570,32, ini menunjukkan bahwa klaster-klaster tersebut memiliki dispersi yang cukup baik, meskipun harus dilihat dalam konteks dataset yang digunakan.
3. Davies-Bouldin Index: Nilai ini adalah 0,46. Indeks ini mengukur rata-rata kesamaan antara setiap klaster dengan klaster yang paling mirip dengannya. Nilai Davies-Bouldin berkisar antara 0 hingga tak terhingga, di mana nilai yang lebih rendah menunjukkan klaster-klaster yang lebih baik. Dalam hal ini, nilai 0,46 menunjukkan bahwa klaster-klaster yang dihasilkan memiliki pemisahan yang relatif baik, meskipun ada ruang untuk perbaikan.

4. KESIMPULAN

Penelitian ini berhasil mengimplementasikan algoritma BIRCH dalam klasterisasi data biaya hidup di beberapa kota di seluruh dunia. Hasil klasterisasi mengungkapkan bahwa kota-kota dengan biaya hidup yang tinggi memiliki kecenderungan harga sewa dan bahan makanan yang lebih tinggi, sementara daya beli lokal tidak selalu berbanding lurus dengan tingginya biaya hidup. Algoritma BIRCH terbukti efektif dalam menangani dataset yang besar serta mendeteksi outlier, yang ditunjukkan dengan hasil evaluasi klaster yang cukup baik melalui Silhouette Score, Calinski-Harabasz Index, dan Davies-Bouldin Index. Visualisasi hasil menggunakan PCA memungkinkan pemetaan klaster yang jelas dan mudah dipahami. Penelitian ini memberikan wawasan penting bagi perencana kebijakan ekonomi kota dalam memahami faktor-faktor yang mempengaruhi biaya hidup dan dapat menjadi acuan dalam pengambilan keputusan terkait pembangunan infrastruktur dan kesejahteraan ekonomi kota.

5. DAFTAR PUSTAKA

S. E. A. Buananta, M. A. Ahmad, J. Mahmood, and P. Paradise, "Identification of evaluation results in e-banking services transaction for product recommendation using the BIRCH and Davies Bouldin index method," Jurnal Infotel, vol. 16, no. 2, pp. 427–440, 2024.

I. Y. Beti and H. Juliansa, "Penerapan normalisasi data metode decimal scaling dan metode k-means dalam mengelompokkan kasus demam berdarah," KLIK: Kajian Ilmiah Informatika dan Komputer, vol. 4, no. 6, pp. 2928–2936, 2024.

J. Laurenso, D. Jiustian, F. Fernando, V. Suhandi, and T. H. Rochadiani, "Implementation of K-means, hierarchical, and BIRCH clustering algorithms to determine marketing targets for vape sales in Indonesia," Journal of Applied Informatics and Computing, vol. 8, no. 1, pp. 62–70, 2024.

M. A. Shauma, Y. Purwanto, and A. Novianty, "Deteksi anomali trafik menggunakan algoritma birch dan dbscan pada streaming traffic," eProceedings of Engineering, vol. 3, no. 3, 2016.

S. Shevira, I. M. A. D. Suarjaya, and P. W. Buana, "Pengaruh kombinasi dan urutan pre-processing pada tweets bahasa indonesia," JITTER-Jurnal Ilmiah Teknologi dan Komputer, vol. 3, no. 2, 2022.

A. Athifaturrofifah, R. Goejantoro, and D. Yuniarti, "Perbandingan pengelompokan k-means dan k-medoids pada data potensi kebakaran hutan/lahan berdasarkan persebaran titik panas," Eksponensial, vol. 10, no. 2, pp. 143–152, 2020.

G. Sumantri, M. D. Novianto, and P. P. Prihastuti, "Implementasi fuzzy c-means dalam pengelompokan provinsi di Indonesia untuk pemerataan kualitas pendidikan," in Prosiding Seminar Pendidikan Matematika dan Matematika, vol. 8, 2023.

F. Ramadhani, "Cluster big data dengan balance intterative reducing clustering using hierarchies (BIRCH)," Ph.D. dissertation, Universitas Sumatera Utara, 2019.

M. A. Shauma, Y. Purwanto, and A. Novianty, "Deteksi anomali trafik menggunakan algoritma birch dan dbscan pada streaming traffic," eProceedings of Engineering, vol. 3, no. 3, 2016.

A. N. Syahrudin and T. Kurniawan, "Input dan output pada bahasa pemrograman python," Jurnal Dasar Pemrograman Python STMIK, vol. 20, pp. 1–7, 2018.

M. U. Munawar and Y. U. Riadi, "Sistem pendeteksi berita hoax di media sosial dengan teknik data mining scikit learn," Jurnal Ilmu Komputer, vol. 4, no. 2, pp. 173–179, 2019.

I. Albanna and R. T. H. Laksono, "Implementasi pandas data frame sebagai agregasi dan tabulasi penyajian data luaran survei kepuasan pengguna proses pembelajaran dalam pendidikan tinggi," in Prosiding Seminar Nasional Sains dan Teknologi Terapan, Nov. 2022.

Y. Galahartlambang, T. Khotiah, and J. Jumain, "Visualisasi data dari dataset COVID-19 menggunakan pemrograman python," Jurnal Ilmiah Intech: Information Technology Journal of UMUS, vol. 3, no. 1, pp. 58–64, 2021.

A. R. Rizalde and H. A. "Comparison of K-means, BIRCH and hierarchical clustering algorithms in clustering OCD symptom data," Journal of Engineering, Data Technology and Computer Science, pp. 102–108, 2024.

I. D. Amirulloh, "Komparasi algoritma clustering data media online pada proses bisnis," Jurnal IKRA-ITH Informatika, pp. 1–8, 2018.

S. E. A. Buananta, "Identification of evaluation results in e-banking services transaction for product recommendation using the BIRCH and Davies Bouldin index method," pp. 427–440, 2024..