

KLASTERISASI PENERIMAAN RETRIBUSI PELAYANAN PASAR MENGUNAKAN METODE K-PROTOTYPE

Dian Ahkam Sani ¹, Moh. Zoqi Sarwani ², Jamaluddin Fahmi ³

^{1, 2, 3} Informatika, Teknologi Informasi, Universitas Merdeka Pasuruan
dian.ahkam@unmerpas.ac.id

ABSTRAK

Pasar tradisional merupakan salah satu pusat kegiatan ekonomi yang penting di Kota Pasuruan. Bagi pedagang, pemilihan tempat berjualan yang strategis dan sesuai dengan kebutuhan pasar merupakan faktor krusial untuk meningkatkan daya jual dan keuntungan. Namun, proses pemilihan tempat berjualan yang optimal seringkali sulit dilakukan karena banyaknya variabel yang perlu dipertimbangkan. Penelitian ini melakukan klasterisasi penerimaan retribusi pelayanan pasar yang ada di kota pasuruan. data yang terkumpul sebanyak 72 data dengan 5 parameter. Penelitian ini menggunakan metode k-prototype sebagai klasterisasinya dan elbow untuk mencari jumlah klaster optimal. Hasil yang didapatkan sebanyak 3 klaster dimana klaster 0 adalah jenis pasar yang penerimaan retribusi pelayanannya rendah sebanyak 23 data, klaster 1 adalah jenis pasar yang jenis penerimaan retribusi pelayanannya tinggi sebanyak 18, sedangkan klaster 2 adalah jenis pasar yang jenis penerimaan retribusi pelayanannya sedang sebanyak 31. Dapat disimpulkan bahwa klaster 0 adalah jenis pasar yang penerimaan retribusi pelayanannya sedikit dilihat dari luas tempat dan harga sewanya, klaster 1 adalah jenis pasar yang jenis penerimaan retribusi pelayanannya kemungkinan banyak dibuktikan dari luas tempat dan juga harga sewanya, sedangkan klaster 2 adalah jenis pasar yang jenis penerimaan retribusi pelayanannya sedang karena luas tempatnya dan juga harga sewa dari tempat tersebut.

Keyword : Pasar Tradisional, Retribusi Pelayanan Pasar, Klaster, K-Prototype, Elbow

1. PENDAHULUAN

Pasar sebagai salah satu pusat kegiatan ekonomi dan perdagangan memiliki peran strategis dalam memenuhi kebutuhan masyarakat. Kota Pasuruan sebagai salah satu kota di Indonesia memiliki sejumlah pasar tradisional yang menjadi pusat transaksi jual beli dan pertemuan antara pedagang dan konsumen. Pemerintah daerah harus memastikan layanan pasar yang optimal untuk mendukung pertumbuhan ekonomi lokal dan kesejahteraan masyarakat.

Retribusi pelayanan pasar merupakan sumber pendapatan utama bagi pemerintah daerah dalam menjalankan fungsinya dalam pengelolaan dan pengembangan pasar. Pendapatan dari retribusi ini dapat digunakan untuk membiayai pemeliharaan, pengembangan infrastruktur, serta peningkatan pelayanan bagi pedagang dan pengunjung pasar. Oleh karena itu, pemahaman mendalam mengenai karakteristik retribusi pelayanan pasar menjadi sangat penting[1]. Pola retribusi pelayanan dapat beragam antara pasar-pasar yang berbeda, tergantung pada faktor-faktor seperti lokasi, ukuran pasar, fasilitas yang disediakan, serta aktivitas perdagangan yang terjadi di dalamnya.

Klasterisasi, sebagai metode analisis data, dapat digunakan untuk mengelompokkan data pasar menjadi kelompok-kelompok dengan karakteristik data serupa[2], [3]. Dalam konteks penerimaan retribusi pelayanan, klasterisasi dapat membantu mengidentifikasi pola-pola pengeluaran retribusi yang berbeda antara pasar-pasar yang berbeda. Dengan demikian, informasi yang dihasilkan dari analisis klasterisasi dapat menjadi dasar untuk

pengambilan keputusan strategis dalam pengelolaan retribusi pelayanan pasar.

Beberapa penelitian terkait klasterisasi pernah dilakukan diantaranya oleh [4] dimana pada penelitian ini menggunakan metode K-Means untuk mengelompokkan siswa berdasarkan proses belajar dan *silhouette coefficient* sebagai metode validasinya mendapatkan *score* 0.69 dengan jumlah *cluster* terbanyak 3. Penelitian selanjutnya [5] adalah untuk melakukan analisis terhadap kinerja algoritma K-Means dan algoritma K-Medoids dalam melakukan proses clustering. Clustering akan dikelompokkan menjadi 4 cluster yaitu: tingkat kinerja sangat baik, tingkat kinerja baik, tingkat kinerja cukup dan tingkat kinerja kurang baik. Sedangkan penelitian yang berkaitan dengan data campuran dilakukan oleh [6] dimana pada penelitian ini menggunakan 18 atribut dengan tipe data yang berbeda-beda menggunakan metode *K-Prototype*. Kualitas pengelompokan dievaluasi menggunakan uji *kruskal-Wallis* yang cocok untuk data tidak normal. Hasilnya menunjukkan bahwa metode ini menghasilkan pengelompokan yang lebih baik dari metode lain dalam hal keakuratan. Penelitian selanjutnya [7] tentang penerapan *clustering* pada tipe data campuran menggunakan *K-Prototype* pada perusahaan *multifinance* dimana hasil yang didapatkan dengan menggunakan 4 *cluster* menunjukkan tingkat kelancaran 100% dan jumlah rasio gaji terhadap angsuran nasabah 1.3%. Penelitian yang menggunakan pengembangan metode *K-Prototype* dilakukan oleh [8] untuk menganalisis data nasabah asuransi jiwa yang terfokus pada karakteristik dan informasi ketentuan

tanpa pertimbangan angka kematian umum. Data penelitian 1.14 juta dari tahun 2014. Analisa menghasilkan 3 klaster.

Dari hasil pemaparan permasalahan dan didukung dengan penelitian terdahulu di atas maka diambil penelitian dengan judul “Klasterisasi Penerimaan Retribusi Pelayanan Pasar Kota Pasuruan Menggunakan Metode K-Prototype”.

2. TINJAUAN PUSTAKA

2.1. Retribusi Pasar

Berdasarkan Peraturan Daerah Kota Pasuruan Nomor 9 Tahun 2011, tentang retribusi pelayanan pasar kota pasuruan. Pasar adalah fasilitas yang disediakan pemerintah daerah sebagai tempat kegiatan jual beli dan bongkar muat komoditi atau barang dagang. Disebut juga bahwa retribusi pasar itu sendiri adalah pungutan retribusi atas jasa pelayanan penyediaan fasilitas pasar tradisional yang berupa pelataran, los dan atau kios atau bedak yang dikelola Pemerintah Daerah dan khusus disediakan untuk pedagang. Retribusi pasar merupakan salah satu Retribusi Daerah yang termasuk dalam jenis Retribusi Jasa Umum. Karena itu dalam retribusi pasar, prinsip dan sasaran dalam penetapan struktur dan besarnya tarif retribusi pasar didasarkan pada tujuan untuk memperoleh keuntungan yang layak sebagai pengganti biaya pengelolaan, biaya penyelenggaraan, biaya kebersihan dan biaya administrasi.

2.2. Klasterisasi

Klasterisasi atau *Clustering* adalah metode pengelompokan data. Menurut [3] klasterisasi adalah sebuah proses untuk mengelompokkan data ke dalam beberapa kelompok sehingga data dalam satu klaster memiliki tingkat kemiripan yang maksimum dan antar klaster memiliki kemiripan yang minimum. *Clustering* sangat berguna untuk prediksi dan Analisa masalah bisnis tertentu. Misalnya marketing, segmentasi pasar dan pemetaan zonasi wilayah. Syarat dalam melakukan klasterisasi adalah skalabilitas, kemampuan Analisa dalam bentuk data, menemukan kelompok dengan bentuk yang tidak terduga, mampu menangani data yang rusak, sensitivitas terhadap perubahan data, interpretasi dan kegunaan. Kualitas dari hasil klasterisasi bergantung pada metode yang dipakai.

2.3. K- Prototype

Algoritma *k-prototype* termasuk dalam metode *non-hierarchical clustering* yang pertama kali dikemukakan oleh Huang (1998). Algoritma ini merupakan algoritma yang digunakan dalam *clustering* yang tipe datanya terdiri dari data campuran (numerik dan kategorik) dengan menggabungkan perhitungan jarak pada algoritma *k-means* dan *k-modes*. Perhitungan jarak *k-prototype* menggunakan kesamaan antara objek data menggunakan penggabungan persamaan

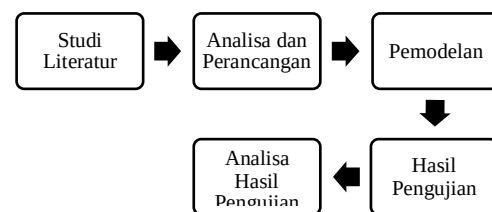
perhitungan jarak *Euclidean* dengan *dissimilarity measure* yang ada dalam *k-modes*. Penggambaran cara kerja algoritma *k-prototype* adalah, misalkan $\{x_{ij}\}$, $i = 1, 2, \dots, n$, $j = 1, 2, \dots, d$ menunjukkan sekumpulan data yang berisi n batas data. Setiap observasi mendeskripsikan variable D . Kemudian untuk *dissimilarity measure* antara dua titik x dan y digunakan pada algoritma *k-prototype* sebagai mana berikut:

$$d_2(X, Y) = \sum_{j=1}^p (x_j - y_j)^2 + \gamma \sum_{j=p+1}^m \delta(x_j, y_j)$$

Dimana $d_2(X, Y)$ adalah ukuran jarak antara objek X dan Y (data campuran), $\sum_{j=1}^p (x_j - y_j)^2$ adalah ukuran jarak untuk data yang bertipe numerik, $\sum_{j=p+1}^m \delta(x_j, y_j)$ adalah ukuran jarak untuk data yang bertipe kategorik dan γ adalah parameter penimbang.

3. METODE PENELITIAN

Bab ini akan memaparkan tentang metode penelitian yang dilakukan pada penelitian ini, yang terdiri dari studi literatur, analisa data, pengujian dengan menggunakan *k-prototype*, hasil pengujian dan analisa hasil pengujian.



Gambar 1. Metode Penelitian

3.1. Studi Literatur

Tahap awal dalam penelitian ini adalah studi literatur berdasarkan topik penelitian yang telah ditentukan. Tahapan pada proses ini meliputi pembuatan konsep mengenai sistem yang akan dibuat dengan melakukan studi literatur berdasarkan topik penelitian yang diambil dari literatur di atas dapat menjadi dasar bagi peneliti untuk merancang, menganalisis dan menentukan jumlah klaster sehingga bisa mendapatkan hasil validasi terbaik. Pengumpulan dataset yang dilakukan adalah dengan cara melakukan survey kepada petugas pasar sebanyak 72 responden.

3.2. Analisa dan Perancangan

Dataset yang didapatkan dalam bentuk survey kepada petugas pasar tradisional Kota Pasuruan sejumlah 72 data. Dari data yang terkumpul terdapat 5 parameter dimana setiap parameter memiliki tipe data kategorikal dan juga angka.

Tabel 1. Tipe data pada parameter.

Column	Non-Null Count	Dtype
nama_pasar	72 non-null	object
letak_tempat	72 non-null	object
jenis_tempat	72 non-null	object
luas_tempat	72 non-null	int64
harga_sewa	72 non-null	int64

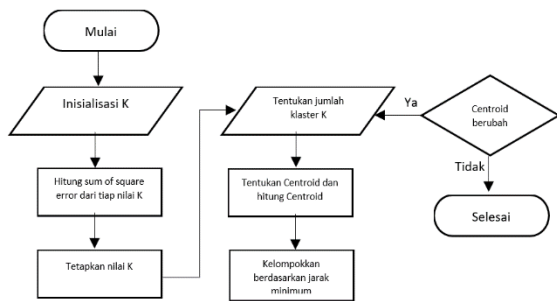
Dari dataset tersebut maka langkah selanjutnya melakukan normalisasi data. Normalisasi data bisa dilakukan jika rentang nilai pada kolom terpaut jauh[9], [10]. Penelitian ini menggunakan *min-max scaler* yang bertujuan untuk mengubah rentang nilai atribut atau fitur menjadi rentang yang lebih kecil antara 0-1[11]. Perhitungan min-max scaler menggunakan rumus sebagai berikut:

$$x' = \frac{x - (x)}{(x) - (x)}$$

dimana X' adalah nilai atribut setelah diubah skala. X adalah nilai atribut asli sebelum diubah skala. $\text{Min}(x)$ adalah nilai minimum dari atribut. $\text{Max}(x)$ adalah nilai maksimum dari atribut.

3.3. Pemodelan

Teknik data mining dengan metode klusterisasi menggunakan metode *k-prototype* dengan kombinasi *elbow* sebagai metode optimasi jumlah cluster sebagaimana gambar 2.

Gambar 2. Flowchart metode *elbow* dan algoritma *k-prototype*.

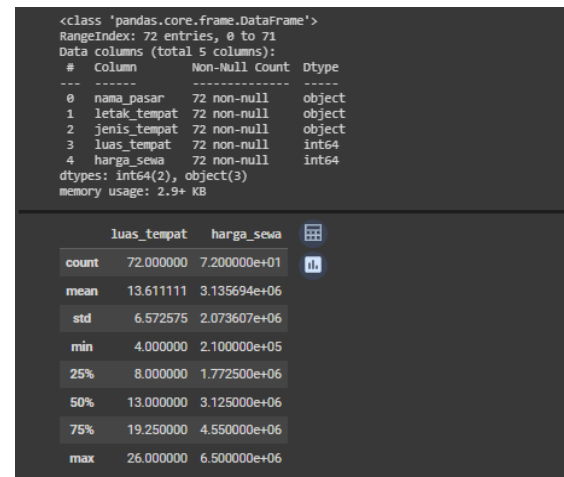
Langkah awal adalah menginisiasi jumlah kluster atau nilai k . selanjutnya adalah menghitung *sum square error* dari tiap-tiap nilai k dan tetapkan nilai K sesuai dengan sikunya. Selanjutnya adalah menghitung nilai centroid sesuai nilai K yang sudah ditetapkan lalu kelompokkan hasilnya. Jika nilai centroid berubah maka atur nilai K , jika tidak berubah maka selesai.

4. HASIL DAN PEMBAHASAN

Pada tahap ini akan dijelaskan hasil dan pembahasan sesuai dengan metode penelitian.

4.1. Analisa dan Perancangan

Pada tahap ini dataset yang telah dikumpulkan dari hasil survey dianalisa dan diolah untuk bisa menentukan klusterisasi penerimaan retribusi pelayanan pasar sesuai dengan atribut yang telah ditentukan. Terdapat beberapa proses yang dilakukan untuk membersihkan data agar bisa membentuk kluster diantaranya adalah mencari data *missing/* hilang, outliers, dan normalisasi data sebagaimana gambar 3.



Gambar 3. Informasi dataset

Dari gambar 3 dapat dilihat bahwa tidak terdapat data yang kosong, namun terdapat beberapa data yang memiliki tipe *object* dan *int64*. Maka langkah selanjutnya adalah mengubah data numerik (*int64*) ke dalam bentuk *float* dan melakukan normalisasi data. Hal ini dikarenakan nominal data pada kolom *luas_tempat* dan *harga_sewa* terpaut jauh. Normalisasi data/ skala fitur pada penelitian ini menggunakan *MinMaxScaler* dimana rentan nilai yang dihasilkan adalah berkisar 0 sampai 1. Hasil dari skala fitur *MinMaxScaler* ditampilkan pada gambar 4.

	nama_pasar	letak_tempat	jenis_tempat	luas_tempat	harga_sewa
0	Pasar Besar	Depan	Kios	0.978874	1.148191
1	Pasar Besar	Tengah	Kios	1.285304	1.391009
2	Pasar Besar	Belakang	Kios	1.438519	1.488136
3	Pasar Besar	Depan	Bedak	0.059584	0.322611
4	Pasar Besar	Tengah	Bedak	0.212799	0.371174

Gambar 4. Hasil *MinMaxScaler()*

a.

4.2. Pemodelan

Setelah proses pembersihan data dilakukan maka langkah selanjutnya adalah mencari nilai kluster yang optimal dengan menggunakan metode *elbow* dengan jumlah iterasi 1 sampai 10.

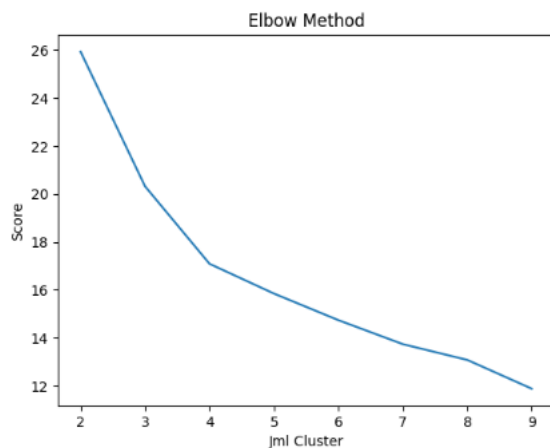
Program Jurnal

```

for k in range (1, 10):
    model = KPrototypes(n_clusters=k, init='Huang',
    verbose=2)
    clusters = model.fit(df, categorical =
    catColumnsPos)
    score.append(model.cost_)
    n_clusters.append(k)
    clusters_assigned.append(clusters)

```

hasil dari algoritma tersebut ditampilkan pada gambar 5.



Gambar 5. Grafik Elbow Method

Gambar di atas menunjukkan terjadinya bentuk siku pada angka 4 yang menandakan bahwa jumlah kluster optimal adalah 4. Setelah mendapat nilai kluster optimal dengan menggunakan *elbow* maka langkah selanjutnya melakukan proses perhitungan dengan menggunakan *k-prototype*.

4.3. Hasil Pengujian

Hasil pengujian dengan menggunakan metode *k-prototype* dengan parameter *n_cluster*= 3, *verbose*= 2, *max_iter*= 20, dan *init*='Huang' menunjukkan iterasi pertama pada run pertama terjadi 8 perubahan dalam mengelompokkan data, dan nilai cost adalah 20.86. Run kedelapan memberikan hasil kluster terbaik dengan nilai cost yang lebih rendah dibandingkan dengan run lainnya. Proses iterasi dari model *k-prototype* ditampilkan pada tabel 2.

Tabel 2. Hasil Pengujian k-prototype

Run	Iteration	Moves	Ncost
1	1/20	8	20.862648385446306
1	2/20	1	20.85708093359511
1	3/20	0	20.85708093359511
2	1/20	0	21.554735409221692
3	1/20	30	20.713406231550486
3	2/20	2	20.68809627412769
3	3/20	0	20.68809627412769
4	1/20	28	20.94941532204891
4	2/20	7	20.579115717660414
4	3/20	3	20.510470166870586

Run	Iteration	Moves	Ncost
4	4/20	0	20.510470166870586
5	1/20	13	20.725071107902664
5	2/20	1	20.719096581632122
5	3/20	0	20.719096581632122
6	1/20	26	20.791549925569687
6	2/20	8	20.47333960803249
6	3/20	1	20.455193496432415
6	4/20	0	20.455193496432415
7	1/20	11	20.827258804933635
7	2/20	7	20.576066772556004
7	3/20	1	20.55537783939929
7	4/20	0	20.55537783939929
8	1/20	9	20.289200906809082
8	2/20	2	20.269348532100203
8	3/20	0	20.269348532100203
9	1/20	18	20.924510982740433
9	2/20	5	20.7959168471884
9	3/20	1	20.777762218511132
9	4/20	0	20.777762218511132
10	1/20	17	20.645720108790364
10	2/20	9	20.31989021419638
10	3/20	1	20.31495302752437
10	4/20	0	20.31495302752437

Secara umum, algoritma ini berusaha mengoptimalkan posisi centroid dan pengelompokan data dalam beberapa iterasi hingga mencapai konvergen. Berikut adalah grafik persebaran data ke dalam 4 kluster.

4.3. Analisa Hasil Pengujian

Analisis kluster dilakukan untuk mendapatkan pola dari hasil kluster yang sudah didapatkan. Terdapat 23 data masuk ke dalam kluster 0, 18 data masuk ke dalam kluster 1 dan 31 data masuk ke dalam kluster 2.

Tabel 3. Analisa hasil kluster.

	label	0	1	2
luas_tempat	median	0.136	0.818	0.500
	mean	0.129	0.837	0.451
harga_sewa	median	0.024	0.920	0.523
	mean	0.120	0.902	0.489

Dengan melihat tabel hasil, dapat diperoleh informasi tentang pola dari tiap-tiap kluster atau disebut dengan segmen. Dapat disimpulkan bahwa kluster 0 adalah jenis pasar yang penerimaan retribusi pelayanannya sedikit dilihat dari luas tempat dan harga sewanya, kluster 1 adalah jenis pasar yang jenis penerimaan retribusi pelayanannya kemungkinan banyak dibuktikan dari luas tempat dan juga harga sewanya, sedangkan kluster 2 adalah jenis pasar yang jenis penerimaan retribusi pelayanannya sedang karena luas tempatnya dan juga harga sewa dari tempat tersebut.

5. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan maka diperoleh beberapa kesimpulan sebagai berikut: Terdapat 3 kluster optimal dengan

menggunakan metode elbow dan juga k-prototype diantaranya klaster 0 adalah jenis pasar yang penerimaan retribusi pelayanannya rendah sebanyak 23 data, klaster 1 adalah jenis pasar yang jenis penerimaan retribusi pelayanannya tinggi sebanyak 18, sedangkan klaster 2 adalah jenis pasar yang jenis penerimaan retribusi pelayanannya sedang sebanyak 31. Klaster 0 adalah jenis pasar yang penerimaan retribusi pelayanannya sedikit dilihat dari luas tempat dan harga sewanya, klaster 1 adalah jenis pasar yang jenis penerimaan retribusi pelayanannya kemungkinan banyak dibuktikan dari luas tempat dan juga harga sewanya, sedangkan klaster 2 adalah jenis pasar yang jenis penerimaan retribusi pelayanannya sedang karena luas tempatnya dan juga harga sewa dari tempat tersebut

DAFTAR PUSTAKA

- [1] Didin Syarifuddin, "Pasar Tradisional Dalam Perspektif Nilai Daya Tarik Wisata," *J. Manaj. Resort dan Leis.*, vol. 15, no. 1, pp. 19–32, 2018.
- [2] Ö. Akay and G. Yüksel, "Hierarchical clustering of mixed variable panel data based on new distance," *Commun. Stat. ...*, 2021, doi: 10.1080/03610918.2019.1588306.
- [3] S. A. Rajagukguk and D. H. Fudholi, "Rich Kid, Poor Kid: Students Clustering using K-Prototype Algorithm," *2022 IEEE 7th Int. ...*, 2022, [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/9971445/>.
- [4] M. Z. Sarwani and D. A. Sani, "Implementasi metode k-means sebagai pengelompokan siswa berdasarkan proses belajar siswa," *Semin. Nas. Sist. Inf. 2018*, pp. 1131–1135, 2018, [Online]. Available: <https://jurnalfti.unmer.ac.id/index.php/senasif/article/download/174/144>.
- [5] A. U. Fitriyadi and A. Kurniawati, "Analisis Algoritma K-Means dan K-Medoids Untuk Clustering Data Kinerja Karyawan Pada Perusahaan Perumahan Nasional," *Kilat*, vol. 10, no. 1, pp. 157–168, 2021, doi: 10.33322/kilat.v10i1.1174.
- [6] Y. Wang, L. Yang, J. Wu, and S. Chen, "An improved K-prototype-based cluster algorithm for mixed educational data mining," *Int. J. ...*, 2022, doi: 10.1504/IJES.2022.127166.
- [7] O. J. Pradhana, D. Saepudin, and Indwiarti., "Penerapan Clustering Pada Tipe Data Campuran Menggunakan K-Prototype Pada Perusahaan Multifinance," *e-Proceding Eng.*, vol. 9, no. Juni, pp. 1996–2007, 2022.
- [8] S. Yin, G. Gan, E. A. Valdez, and J. Vadiveloo, "Applications of clustering with mixed type data in life insurance," *Risks*, 2021, [Online]. Available: <https://www.mdpi.com/2227-9091/9/3/47>.
- [9] D. U. Ozsahin, M. T. Mustapha, A. S. Mubarak, and ..., "Impact of feature scaling on machine learning models for the diagnosis of diabetes," ... *Everything*, 2022, [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/9898687/>.
- [10] Q. Chen, Y. Wang, T. Yang, X. Zhang, and ..., "You only look one-level feature," *Proc. ...*, 2021, [Online]. Available: http://openaccess.thecvf.com/content/CVPR2021/html/Chen_You_Only_Look_One-Level_Feature_CVPR_2021_paper.html.
- [11] B. E. Woodworth, B. Bullins, O. Shamir, and ..., "The min-max complexity of distributed stochastic convex optimization with intermittent communication," ... *Learn. Theory*, 2021, [Online]. Available: <http://proceedings.mlr.press/v134/woodworth21a.html>.
- [12] E. Umargono, J. E. Suseno, and ..., "K-means clustering optimization using the elbow method and early centroid determination based on mean and median formula," *2nd Int. ...*, 2020, [Online]. Available: <https://www.atlantispress.com/proceedings/isstec-19/125944915>.
- [13] M. Cui, "Introduction to the k-means clustering algorithm based on the elbow method," *Accounting, Audit. Financ.*, 2020, [Online]. Available: <https://www.clausiuspress.com/article/592.html>.
- [14] R. Bi, D. Guo, Y. Zhang, R. Huang, L. Lin, and ..., "Outsourced and Privacy-Preserving Collaborative K-Prototype Clustering for Mixed Data via Additive Secret Sharing," *IEEE Internet Things ...*, 2023, [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/10098548/>.
- [15] N. P. Hariani, Kariyam, and S. K. Dini, "Pengelompokan Perguruan Tinggi Negeri Di Indonesia Menggunakan Metode Hirarki Dan K-Medoids Dengan Ukuran Jarak Modifikasi Data Campuran," *Ajie*, vol. 06, no. 20, pp. 87–98, 2022, doi: 10.20885/ajie.vol6.iss2.art5.