

BAB V

SIMPULAN DAN SARAN

5.1 Simpulan

Dari hasil evaluasi *cluster* yang dilakukan menggunakan metode K-means dengan dua metrik evaluasi, yaitu *davies-bouldin index* (DBI) sebesar 0.00178 dan *silhouette score* sebesar 0.9972 menghasilkan 2 *cluster*, dimana *cluster* 1 terdapat 1518 titik dengan rata-rata gaji Rp 7jt dan *cluster* 2 terdapat 1482 titik dengan rata-rata gaji Rp 8jt, dapat ditarik kesimpulan sebagai berikut:

1. *Davies-Bouldin Index* (DBI): Nilai DBI sangat rendah, mendekati nol, yang menunjukkan bahwa klastering yang dihasilkan sangat baik. Ini mengindikasikan bahwa jarak antara *cluster* relatif besar dibandingkan dengan ukuran *cluster* itu sendiri.
2. *Silhouette Score* : Nilai *silhouette score* yang sangat tinggi, mendekati 1, menunjukkan bahwa klastering yang dihasilkan sangat baik. Nilai *silhouette score* yang tinggi menandakan bahwa setiap data berada dekat dengan *cluster* lainnya dan jauh dari *cluster* lain, yang mengindikasikan bahwa klastering tersebut konsisten dan terpisah dengan baik.
3. Jumlah *Cluster* Optimal : Berdasarkan evaluasi yang dilakukan, jumlah *cluster* optimal yang direkomendasikan adalah 2. Hal ini didukung oleh kedua metrik evaluasi yang menunjukkan hasil yang sangat baik untuk klastering dengan 2 *cluster*.

Dengan demikian, dapat disimpulkan bahwa klastering yang dilakukan menggunakan metode K-means dengan 2 *cluster* menghasilkan hasil yang sangat

baik berdasarkan kedua metrik evaluasi yang digunakan.

5.2 Saran

Berdasarkan penelitian yang telah dilakukan menggunakan metode K-Means *clustering* dalam analisis gaji berdasarkan data demografis, terdapat beberapa saran yang dapat diajukan:

1. Metode K-Means *Clustering* telah terbukti efektif dalam mengelompokkan data gaji berdasarkan faktor-faktor demografis. Saran untuk penelitian selanjutnya adalah mempertimbangkan penggunaan algoritma *clustering* lainnya, seperti *Hierarchical Clustering* atau DBSCAN, untuk membandingkan kinerja dan hasil klasteringnya.
2. Meskipun K-Means *Clustering* telah memberikan hasil yang memuaskan, penelitian selanjutnya dapat menggali lebih dalam untuk memahami karakteristik masing-masing *cluster* dengan lebih detail. Ini bisa mencakup analisis perbandingan antar *cluster*, identifikasi faktor-faktor yang paling mempengaruhi pengelompokan gaji dalam setiap *cluster*, atau eksplorasi tren atau pola khusus di dalam *cluster* tertentu. Pengembangan model prediktif untuk memprediksi gaji berdasarkan fitur demografis juga dapat menjadi langkah yang menarik.
3. Hasil penelitian ini dapat membentuk dasar untuk pengembangan sistem pendukung keputusan (*Decision Support System*) yang dapat membantu perusahaan atau organisasi dalam pengelolaan gaji karyawan. Saran untuk peneliti berikutnya adalah mengintegrasikan hasil analisis klastering gaji ke dalam sistem yang dapat memberikan rekomendasi atau strategi yang lebih

efektif dalam penetapan gaji berdasarkan faktor-faktor demografis yang relevan.

DAFTAR PUSTAKA

- Athey, T. L., Liu, T., Pedigo, B. D., & Vogelstein, J. T. (2019). Autogmm: Automatic and hierarchical gaussian mixture modeling in python. arXiv preprint arXiv:1909.02688.
- Anggarawati, S. (2022). Penerapan Algoritma K-Means *Clustering* Untuk Menentukan Linieritas Pekerjaan Alumni Berdasarkan Tracer Study. In Seminar Nasional Sistem Informasi (SENASIF) (Vol. 6, pp. 3265-3281).
- Amoozad Mahdiraji, H., Tavana, M., Mahdiani, P., & Abbasi Kamardi, A. A. (2022). A multi-attribute data mining model for rule extraction and service operations benchmarking. *Benchmarking: An International Journal*, 29(2), 456-495.
- Badan Pusat Statistik. (2022). Berita Resmi Statistik. Jakarta : Badan Pusat Statistik.
- Das, S., Barik, R., & Mukherjee, A. (2020). Salary prediction using regression techniques. *Proceedings of Industry Interactive Innovations in Science, Engineering & Technology (I3SET2K19)*.
- Drl, I. R., Chrisnanto, Y. H., & Umbara, F. R. (2022). Analisis *Cluster* Pada Kelompok Masyarakat Yang Rentan Terhadap Paparan Covid-19 Menggunakan Metode K-Means *Clustering* Dan Visualiasi Dengan Sig. *Informatics and Digital Expert (INDEX)*, 4(2), 61-69.
- Ekaningtyas, R. M. (2020). Persaingan Dan Diskriminasi Upah Gender Di Industri Manufaktur Indonesia. *Ilmiah Ekonomi Dan Bisnis*, 17(2), 168-175.
- Elshawi, R., & Sakr, S. (2020). Automated machine learning: Techniques and frameworks. In *Big Data Management and Analytics: 9th European Summer School, eBISS 2019, Berlin, Germany, June 30–July 5, 2019, Revised Selected Papers 9* (pp. 40-69). Springer International Publishing.
- Faisal, M., & Zamzami, E. M. (2020, June). Comparative analysis of inter-centroid K-means performance using Euclidean distance, canberra distance and manhattan distance. In *Journal of Physics: Conference Series* (Vol. 1566, No. 1, p. 012112). IOP Publishing.
- Firdaus, B. A., Ratnawati, D. E., & Hanggara, B. T. (2021). Klusterisasi Popularitas Artist pada Playlist Today's Top Hits Menggunakan Metode K-Means dengan Integrasi Spotify Web API dan Teknologi Amazon SageMaker. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 5(1), 373-380.
- Glüge, S., Amirian, M., Flumini, D., & Stadelmann, T. (2020, September). How (not) to measure bias in face recognition networks. In *IAPR Workshop on Artificial Neural Networks in Pattern Recognition* (pp. 125-137). Cham: Springer International Publishing.
- Hunter, J., Dale, D., Firing, E., & Droettboom, M. (2020). Tutorials — Matplotlib 3.1.2 Documentation. <https://matplotlib.org/3.1.1/tutorials/index.html>.
- Idehlu, H. A., Sadeghabad, A. J., & Pagheh, B. (2019). Increasing the efficiency and effectiveness of government organizations with the contracts of salary payment on the basis of consequence and outcome. *International Journal of Innovative Technology and Exploring Engineering*, 8(11), 2443-2445.
- Ilbeigipour, S., Albadvi, A., & Noughabi, E. A. (2022). *Cluster-based analysis of*

- COVID-19 cases using self-organizing map neural network and K-means *methods* to improve medical decision-making. *Informatics in Medicine Unlocked*, 32, 101005.
- Jurczyk, T. (2021). *Clustering with Scikit-Learn in Python*. The Programming Historian.
- Karna, A., & Gibert, K. (2022). Automatic identification of the number of *Clusters* in hierarchical *Clustering*. *Neural Computing and Applications*, 34(1), 119-134.
- Liao, Q. Z., Xue, L., Lei, G., Liu, X., Sun, S. Y., & Patil, S. (2022). Statistical prediction of waterflooding performance by K-means *Clustering* and empirical modeling. *Petroleum Science*, 19(3), 1139-1152.
- Martin, I., Mariello, A., Battiti, R., & Hernández, J. A. (2018). Salary prediction in the it job market with few high-dimensional samples: A spanish case study. *International Journal of Computational Intelligence Systems*, 11(1), 1192-1209.
- Munti, N.Y.S., Nurcahyo, G.W., & Santony, J. (2018). Analisis Dan Penerapan Data Mining Untuk Menentukan Gaji Karyawan Tetap Dan Karyawan Kontrak Menggunakan Algoritma K-means *Clustering* (Studi Kasus Di PT Indomex Dwijaya Lestari). *Jurnal Informatika Teknologi Informasi*, 1(1), ISSN 2620-6153.
- Papastathopoulos, A., Ahmad, S. Z., Al Sabri, N., & Kaminakis, K. (2020). Demographic analysis of residents' support for tourism development in the UAE: A Bayesian structural equation modeling multigroup approach. *Journal of Travel Research*, 59(6), 1119-1139.
- Puspasari, B. D., Damayanti, L. L., Pramono, A., & Darmawan, A. K. (2021, October). Implementation K-means *Clustering method* in job recommendation system. In *2021 7th International Conference on Electrical, Electronics and Information Engineering (ICEEIE)* (pp. 1-6). IEEE.
- Ramadhani, R., & Firmansyah, Y. (2023). Dampak Keberagaman Demografis Tenaga Kerja Terhadap Kinerja Karyawan Di Pt Xyz. *Ultima Management: Jurnal Ilmu Manajemen*, 15(1), 100-123.
- Ros, F., Riad, R., & Guillaume, S. (2023). PDBI: A partitioning *Davies-Bouldin Index* for *Clustering* evaluation. *Neurocomputing*, 528, 178-199.
- Rybak, O. (2019). The Main Properties Of Salary And Its Regulation.
- Sinaga, K. P., & Yang, M. S. (2020). Unsupervised K-means *Clustering* algorithm. *IEEE access*, 8, 80716-80727.
- Setiawan, Z., Fajar, M., Priyatno, A. M., Putri, A. Y. P., Aryuni, M., Yuliyanti, S., & Wijaya, A. (2023). *Buku Ajar Data Mining*. PT. Sonpedia Publishing Indonesia.
- Syaputri, D., Noprita, P. H., & Romelah, S. (2021). Implemnetation of K-means Algorithm for Economic Distribution *Clustering* Base on Demographics of Population. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 1(1), 1-6.
- Tim pengembangan pandas. *Pandas.DataFrame* — pandas 2.1.1 documentation. 2023, https://pandas.pydata.org/docs/getting_started/index.html
- Twang, A. H. (2022). High-Impact Learning Experiences and Post-Graduate

- Outcomes: Exploring the Influence on Employment, Continuing Education and Salary. *The SUNY Journal of the Scholarship of Engagement: JoSE*, 2(1), 2.
- Vadyala, S. R., Betgeri, S. N., Sherer, E. A., & Amritphale, A. (2021). Prediction of the number of COVID-19 confirmed cases based on K-means-LSTM. *Array*, 11, 100085.
- Vergani, A. A., & Binaghi, E. (2018, July). A soft davies-bouldin separation measure. In *2018 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)* (pp. 1-8). IEEE.
- Zhang, G., Zhang, C., & Zhang, H. (2018). Improved K-means algorithm based on density Canopy. *Knowledge-based systems*, 145, 289-297.
- Zhang, X., Huang, J., Yang, Y., He, X., Liu, R., & Zhong, N. (2019, December). Applying python in brain science education. In *2019 International Joint Conference on Information, Media and Engineering (IJCIME)* (pp. 396-400). IEEE.

LAMPIRAN

Berikut di bawah ini merupakan tabel kode transformasi setiap atribut :

Tabel Transformasi atribut Gender

Gender	kode
Male	1
Female	0

Tabel Transformasi atribut education_level

education_level	kode
sma/smk/stm	4
d3	3
Sarjana/d4	2
Magister/S2	1
Doktor/S3	0

Tabel Transformasi atribut career_level

career_level	kode
pegawai	3
manajer/asisten manajer	2
supervisor/koordinator	4
lulusan baru/pengalaman kerja kurang dari 1 tahun'	1
ceo/gm/direktur/manajer senior	0

Tabel Transformasi atribut location

location	kode	location	kode	location	kode
jakarta pusat	39	palu	91	jakarta utara	43
jakarta raya	40	aceh	0	jakarta barat	38
surabaya	117	bali	3	bekasi	15
yogyakarta	129	tasikmalaya	123	bandung	6
tangerang	120	sidoarjo	108	bogor	20
kalimantan b	51	banten	11	sumatera utara	116
depok	34	jakarta selatan	41	bengkulu	16
jambi	44	bandung	2	jakarta timur	42
banjarasin	10	purwakarta	101	medan	80
pontianak	99	batam	14	jawa tengah	46
dumai	35	bangka	7	malang	72
kutai timur	66	makassar	71	kalimantan selatan	52
jepara	50	surakarta	118	NTT	86
sumatera sel	115	papua	92	denpasar	33
bontang	21	jawa timur	47	jayapura	48
karawang	56	muara enim	82	cikarang	25
bitung	19	semarang	106	pasuruan	95
palembang	89	cikupa	26	sulawesi selatan	112
jawa barat	45	kalimantan timur	54	balikpapan	4
murung	83	raya	84	pekalongan	104
riau	37	gresik	109	sleman	28

cilegon	105	samarinda	97	pekanbaru	59
klaten	49	jember	111	sulawesi barat	122
tarakan	124	tegal	85	NTB	126
timika	78	maros	69	madiun	53
kalimantan t	24	cibinong	113	sulawesi tengah	110
sukabumi	67	citeureup	32	bandar lampung	5
bangka bel	8	pematangsiantar	98	cimahi	30
cirebon	31	banda aceh	61	mataram	79
ternate	125	magelang	70	bintan	18
serang	107	ungaran	128	padang	87
kotawaringin t	63	maluku	73	palopo	90
sulawesi teng	114	kotawaringin b	62	palangkaraya	88
bantul	12	mojokerto	81	brebes	22
banjar	9	pare-pare	94	monkowari	77
ambon	1	lombok	68	ubud	127
cileungsi	29	banyuwangi	13	manado	76
kediri	57	tanjung pinang	121	kutai barat	65
prabumulih	100	gorontalo	36	rangkasbitung	103
purwokerto	102	kendari	58	tanah bambu	119
kalimantan u	55	nunukan	84	maluku utara	74
bima	17	mamuju	75	kudus	64
cianjur	23	klungkung	60	papua barat	93
cilacap	27				

Tabel Transformasi atribut company_industry

Company_industry	kode
komputer/teknik informatika	18
asuransi	2
manufaktur/produksi	27
retail/merchandise	46
makanan & minuman/katering/restoran	25
lainnya	21
transportasi/logistik	51
umum & grosir	53
konstruksi/bangunan/teknik	19
perbankan/pelayanan keuangan	38
elektrikal & elektronik	7
manajemen/konsulting hr	26
automobil/mesin tambahan automotif/kendaraan	3
perawatan/kecantikan/fitnes	37
konsultasi	20
properti/real estate	45
bahan kimia/pupuk/pestisida	4
pendidikan	36
pelayanan arsitek/desain interior	33
kesehatan/medis	17
akunting / audit / layanan pajak	1
hukum/legal	10
call center/it-enabled services/bpo	6
pameran/manajemen acara/pikp	32

periklanan/marketing/promosi/hubungan masyarakat	40
industri berat/mesin/peralatan	12
Pertambangan	42
produk konsumen/barang konsumen yang bergerak cepat	44
Telekomunikasi	49
hiburan/media	8
lingkungan/kesehatan/keamanan	23
inyak/gas/petroleum	28
agrikultural/perkebunan/peternakan unggas/perikanan	0
tekstil/garment	48
hotel/pariwisata	9
Olahraga	29
seni/desain/fashion	47
permata/perhiasan	41
layanan umum/tenaga penggerak	22
pelayanan perbaikan & pemeliharaan'	34
percetakan/penerbitan	39
organisasi nirlaba/pelayanan sosial/lsm	30
luar angkasa/aviasi/pesawat terbang	24
Pakaian	31
kayu/fiber/kertas	14
kelautan/aquakultur	16
travel/pariwisata	52
bioteknologi/farmasi/riset klinik	5
ilmu pengetahuan & teknologi	11
pemerintahan/pertahanan'	35
polymer/plastik/karet/ban	43
keamanan/penegak hukum	15
Tembakau	50
jual beli saham/sekuritas	13

Library Python

```
import matplotlib.pyplot as plt
import pandas as pd
import pylab as pl
import numpy as np
%matplotlib inline
import seaborn as sns
from sklearn.preprocessing import MinMaxScaler
from sklearn.cluster import KMeans
from sklearn.metrics import davies_bouldin_score
from sklearn.metrics import silhouette_score
```