

PENENTUAN *SISTER CITY* UNTUK DIAGRAM TIMBANG DI NUSA TENGGARA TIMUR DENGAN ALGORITMA *K-MEANS*

Putu Dita Pickupana¹, Putu Hadi Purnama Jati², Muhamad Sukin³

¹Fungsi Statistik Distribusi, Badan Pusat Statistik Provinsi NTT, Indonesia

²Fungsi IPDS, Badan Pusat Statistik Provinsi NTT, Indonesia

³Fungsi Statistik Sosial, Badan Pusat Statistik Provinsi NTT, Indonesia

✉korespondensi author: pickupan@bps.go.id, putuhadi@bps.go.id, sukin@bps.go.id

Abstract

The Consumer Price Index (IHK) is a value that calculates changes in the weighted average price of goods and services consumed by households and serves as the basis for the BPS-Statistics to calculate inflation. Weighting data from the BPS-Statistics cost of living survey (SBH) is one of the components required to explain and demonstrate the dynamics of the IHK. SBH is held in several cities due to the limited resources to conduct this survey. As a solution, the sister city approach is adopted by BPS-Statistics to estimate the consumer price index for cities that are not part of the cost-of-living survey domain. The sister city approach uses weighting data from a city that held SBH with similar consumption patterns and is located geographically close to each other. Although the appointment of a sister city went through several procedures, there was no existing method to measure how similar a city is to another city based on the sister city definition. In this paper, we will use machine learning to analyze the similarity of cities in Nusa Tenggara Timur based on their consumption patterns, and as a result, the decision to appoint a sister city will be more accurate. Machine learning is a field of artificial intelligence (AI) and computer science that uses data and algorithms to mimic how people learn and progressively increase its accuracy. Machine learning methods will support the sister city approach with scientific reasoning to produce more accurate inflation. The result of the clustering based on the elbow method for K-means shows that Kupang city has a unique characteristic which means there is no similarity with the other cities in Nusa Tenggara Timur. However, other cities grouped into two cluster where the two inflation cities (Maumere dan Waingapu) are not in the same cluster.

Keyword: *Inflation Rate, Sister City, Basket of Goods, K-Means Algorithm*

1. Pendahuluan

Inflasi adalah salah satu indikator perekonomian paling penting dalam suatu negara dan oleh karena itu, masyarakat sering mengaitkan

inflasi dengan gejolak sosial, politik dan ekonomi di dalam maupun di luar negeri. Tingkat inflasi yang rendah dan stabil adalah prasyarat untuk pertumbuhan ekonomi yang berkesinambungan dan pada akhirnya

mewujudkan kesejahteraan masyarakat. Inflasi yang tinggi dan tidak stabil akan memberikan efek negatif pada kondisi sosial ekonomi masyarakat (Bank Indonesia, n.d.).

Pengendalian tingkat inflasi atau menjaga kestabilan harga menjadi sangat krusial dan selalu menjadi perhatian pemerintah karena beberapa alasan. Pertama, inflasi bisa membuat distribusi pendapatan tidak seimbang. Kedua, inflasi bisa mengurangi tabungan domestik yang merupakan sumber dana investasi negara-negara berkembang. Ketiga, defisit neraca perdagangan dan utang luar negeri yang besar bisa diakibatkan oleh inflasi. Dan terakhir, inflasi juga bisa menyebabkan keadaan politik yang tidak stabil (Kholis & Ferdian, 2019).

Upaya pemerintah untuk mengendalikan inflasi terlihat dari dibentuknya Tim Pemantauan dan Pengendalian Inflasi (TPI) di level pusat sejak tahun 2005 dan dibentuknya Tim Pengendalian Inflasi di level daerah (TPID) mulai tahun 2008. Pada akhirnya TPID memiliki dasar hukum dan pengelolaan berdasarkan Instruksi Menteri Dalam Negeri (Inmendagri) No. 027/1696/SJ tanggal 2 April 2013 tentang “Menjaga Keterjangkauan Barang dan Jasa di Daerah” (Tim Pelaksana Kelompok Kerja Nasional TPID, 2014).

Badan Pusat Statistik (BPS) adalah lembaga yang memiliki tugas untuk melakukan penghitungan inflasi di Indonesia. Dalam menghitung inflasi,

BPS memakai Indeks Harga Konsumen (IHK) sebagai basis perhitungan inflasi dengan mengkalkulasi perubahan rata-rata harga dari barang dan jasa yang dikonsumsi oleh rumah tangga.

Diagram timbang dari Survei Biaya Hidup (SBH) yang dilakukan oleh BPS adalah salah satu komponen yang dibutuhkan untuk menjelaskan dan memperlihatkan dinamika dari IHK. Walaupun TPID terdapat di seluruh wilayah Indonesia, tetapi SBH hanya dilaksanakan di 90 kota oleh karena biaya pelaksanaannya cukup besar dan membutuhkan sumber daya manusia yang banyak untuk melakukan survei ini.

Pendekatan yang diadopsi oleh BPS untuk tetap memunculkan nilai inflasi pada kabupaten/kota yang tidak melakukan kegiatan SBH adalah pendekatan *sister city* (Badan Pusat Statistik, 2015). Pendekatan ini menggunakan diagram timbang dari suatu kota pelaksana SBH yang memiliki pola konsumsi sama dan letaknya berdekatan secara geografis (Badan Pusat Statistik, 2017). Variabel lain yang juga dipertimbangkan dalam pendekatan *sister city* adalah Pendapatan Domestik Regional Bruto (PDRB) dan populasi dari *sister city*.

Meskipun penentuan *sister city* dengan berbagai variabel kemiripan selama ini dipakai oleh BPS, namun belum ada sebuah metode yang dikembangkan secara standar untuk mengukur seberapa mirip sebuah kota yang tergabung dalam *sister city*.

Pengelompokan *sister city* dengan cara yang ilmiah menjadi penting karena metode yang dapat mengukur kesamaan *sister city* akan membuat estimasi inflasi di suatu daerah mendekati nilai sebenarnya.

Dalam tulisan ini, analisis yang digunakan adalah pengelompokan *K-Means* dalam *machine learning* untuk menganalisis kesamaan antar kabupaten dan kota di Nusa Tenggara Timur sehingga keputusan mengelompokkan berdasarkan *sister city* menjadi lebih akurat. *Machine learning* adalah bidang kecerdasan buatan (AI) dan ilmu komputer yang menggunakan data, algoritma untuk meniru cara manusia belajar, dan semakin meningkatkan akurasi (Mintz & Brodie, 2019). Efek *machine learning* juga telah dirasakan secara luas di seluruh ilmu komputer dan industri yang berkaitan dengan pemakaian data yang intensif.

Selain itu, *machine learning* juga memiliki efek yang sama pada seluruh ilmu pengetahuan mulai dari ilmu biologi, kosmologi, hingga ilmu sosial karena *machine learning* dikembangkan untuk menganalisa data eksperimen dengan jumlah yang besar. Metode *machine learning* akan dipakai untuk mendukung pendekatan *sister city* dengan ilmiah agar dapat menghasilkan inflasi yang lebih akurat.

Karena penelitian ini mengarah pada *unsupervised learning* dimana pengelompokan akan dilakukan tanpa *training data* atau label yang sudah

ditentukan, maka nantinya akan digunakan pendekatan *clustering* untuk penelitian ini. Algoritma yang akan digunakan adalah algoritma *K-Means* yang merupakan salah satu metode populer dalam *unsupervised learning*.

2. Metodologi

Bahan dan Sumber Data

Penelitian ini menggunakan data pengeluaran rumah tangga dari hasil Survei Sosial Ekonomi Nasional (Susenas) Maret 2020 oleh Badan Pusat Statistik yang telah diolah. Susenas merupakan salah satu survei rutin BPS yang menjadi sumber data utama untuk kegiatan pembangunan bidang sosial dan ekonomi. Hasil dari Susenas menjadi rujukan berbagai pihak baik pemerintah dan swasta dan juga menjadi rujukan utama memantau perkembangan pencapaian Sustainable Development Goals (SDGs) setiap tahunnya hingga tahun 2030 (buku saku Susenas).

Susenas Maret 2020 merupakan salah satu survei BPS dengan jumlah sampel terbesar, pada Provinsi NTT sendiri mencapai 12.630 rumah tangga dan mampu menghasilkan estimasi hingga level Kabupaten/Kota. Susenas dalam kegiatan surveinya mewawancarai rumah tangga yang menjadi sampel secara *door-to-door* terkait dengan semua pengeluaran makanan dan non makanan. Item pertanyaan untuk pengeluaran rumah tangga sampel Susenas cukup lengkap

sehingga setiap komoditi yang tercakup dalam basket komoditi penyusun IHK, ditanyakan juga pada rumah tangga sampel.

Cakupan sampel Susenas tergolong besar dan meliputi seluruh kabupaten/kota di Provinsi NTT. Sampel dibagi lagi masing-masing menjadi daerah perdesaan dan perkotaan. Dengan jumlah sampel yang besar, diharapkan data Susenas akan mampu memberikan gambaran pola konsumsi dan keterbandingan pengeluaran yang lebih jelas antara satu daerah dengan daerah lainnya.

Berdasarkan pertimbangan item pertanyaan dan cakupan sampel seluruh wilayah di Provinsi NTT, maka Susenas akan digunakan sebagai sumber data utama dalam penelitian ini. Data yang digunakan terdiri atas 14 variabel pengeluaran perkapita makanan dan 6 variabel pengeluaran perkapita non-makanan. Pengeluaran ini mirip dengan 11 kelompok pengeluaran dalam kelompok inflasi.

Pengeluaran makanan terdiri atas kelompok padi-padian, umbi-umbian, ikan, daging, telur dan susu, sayur-sayuran, kacang-kacangan, buah-buahan, minyak dan kelapa, bahan minuman, bumbu-bumbuan, bahan makanan lainnya, makanan minuman jadi, serta rokok dan tembakau. Sedangkan pengeluaran non-makanan terdiri atas kelompok perumahan dan fasilitas rumah tangga, aneka barang dan jasa, pakaian alas kaki dan tutup kepala, barang tahan lama, pajak

pungutan dan asuransi, serta keperluan pesta dan upacara/kenduri. Untuk kelompok pengeluaran pada data inflasi/indeks harga konsumen terdiri atas kelompok makanan minuman dan tembakau, pakaian dan alas kaki, perumahan air listrik dan bahan bakar lainnya, perlengkapan peralatan dan pemeliharaan rutin rumah tangga, kesehatan, transportasi, informasi komunikasi dan jasa keuangan, rekreasi olahraga dan budaya, pendidikan, penyediaan makanan dan minuman atau restoran, serta kelompok perawatan pribadi dan jasa lainnya.

Selain dari data Susenas, penelitian ini juga menggunakan data Produk Domestik Regional Bruto (PDRB) per Kapita Atas Dasar Harga Berlaku tahun 2020. PDRB per kapita sendiri adalah PDRB atas dasar harga pasar yang dibagi dengan jumlah penduduk pertengahan tahun (BPS, 2020). Seperti diketahui bahwa PDRB per kapita memiliki hubungan yang signifikan dengan pola konsumsi dan pendapatan masyarakat, dan cenderung lebih kuat di negara-negara berpendapatan rendah dan menengah (Diacon & Maha, 2015).

Metode Analisis Data

Berdasarkan literatur, *machine learning* memiliki beberapa definisi. Menurut Roihan, dkk (2020), *machine learning* didefinisikan sebagai aplikasi komputer dan algoritma matematika yang mempelajari data yang ada untuk menghasilkan prediksi data masa yang akan datang. Proses pembelajaran itu

melalui dua tahapan yaitu Latihan (*training*) dan pengujian (*testing*).

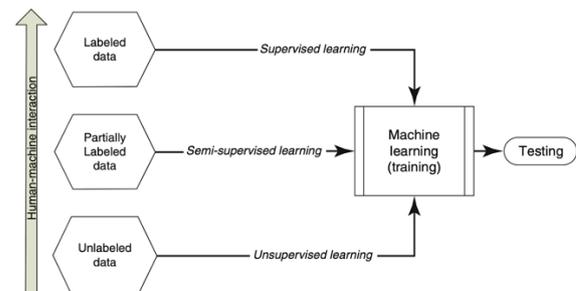
Tom Mitchel (1959) memakai kamus ilmu komputer untuk mendefinisikan *machine learning* sebagai sebuah program yang yang belajar dari *experience* (E) sehubungan dengan beberapa *task* (T) dan pengukuran *performance* (P). Menurut Ethem Alpaydin (2020) *machine learning* didefinisikan sebagai "pemrograman komputer untuk mengoptimalkan kriteria kinerja menggunakan data contoh atau pengalaman masa lalu".

Machine learning bisa dibagi berdasarkan jenis pelabelan datanya. *Supervised learning*, salah satu jenis *machine learning*, memerlukan data yang sudah memiliki jawaban (label) untuk membangun sebuah model yang akurasiya bisa ditingkatkan dari waktu ke waktu seiring dengan bertambahnya data yang diolah. Contoh dari *supervised learning* adalah klasifikasi dan regresi.

Jenis *machine learning* berikutnya adalah *unsupervised learning*, dimana hanya sample input yang diberikan ke sistem pembelajaran, contohnya adalah *clustering* dimana kita berusaha untuk mendapatkan kelompok data secara alami sehingga setiap anggota kluster memiliki varian yang kecil dan antar kelompok kluster memiliki varian yang tinggi. Tidak ada label yang disediakan dari data awal.

Terakhir, *Semi-supervised learning* dimana merupakan kombinasi

antara *supervised learning* dan *unsupervised learning* dimana data sebagian sudah memiliki label dan digunakan untuk menduga data yang belum berlabel.



Gambar 1. Kategori Algoritma *Machine Learning*

Penelitian ini juga menggunakan algoritma *K-means*, salah satu metode yang banyak digunakan pada *cluster analysis* dan juga biasa digunakan pada *data mining* dan pengenalan pola. Fondasi dari algoritma *K-means* adalah meminimalisasi *cluster performance index*, *square-error*, dan *error criterion* (Li & Wu, 2012).

Windarto (2017) menjelaskan bahwa algoritma yang digunakan dalam analisis *K-Means* merupakan model *centroid*. *Centroid* disini berarti titik tengah sebuah *cluster* yang dibentuk. Sebuah obyek akan masuk ke suatu *cluster* dengan *centroid* terdekat atau memiliki jarak terpendek dengan obyek. Berikut adalah *flowcharts* atau alur iterasi pembentukan *cluster* dalam analisis *K-Means* menurut Hermawati (2013):



Gambar 2. Alur (*Flow Chart*) Iterasi Pembentukan *Cluster* dalam Analisis *K-Means*

Elbow Method

Elbow Method adalah suatu metode yang dipakai untuk menentukan jumlah *cluster* terbaik dengan melihat persentase hasil perbandingan antara jumlah *cluster* yang membentuk siku pada suatu titik. Ide dan gagasan dari metode ini adalah memilih nilai *cluster* lalu menambah nilai *cluster* tersebut menjadi model data dalam penentuan *cluster* terbaik (Muningsih, 2017)

Untuk mendapatkan perbandingan, metode ini menghitung Sum of Square Error (SSE) dari masing-masing nilai *cluster*. Semakin besar nilai jumlah *cluster* K maka nilai SSE akan semakin kecil. Rumus SSE pada K-means

sebagai berikut :

$$SSE = \sum_{K=1}^K \sum_{xi \in Sk} \|Xi - Ck\|_2^2$$

Setelah itu dilihat nilai K yang mengalami penurunan paling besar dan berikutnya hasil dari nilai K akan turun secara perlahan sampai nilainya menjadi stabil. Algoritma dari Elbow Method dalam menentukan nilai K pada K-means adalah sebagai berikut :

1. Mulai
2. Inisialisasi awal nilai K
3. Naikkan nilai K
4. Hitung hasil SSE dari tiap nilai K
5. Melihat hasil SSE dari nilai K yang menurun drastic
6. Tetapkan nilai K yang berbentuk siku
7. Selesai

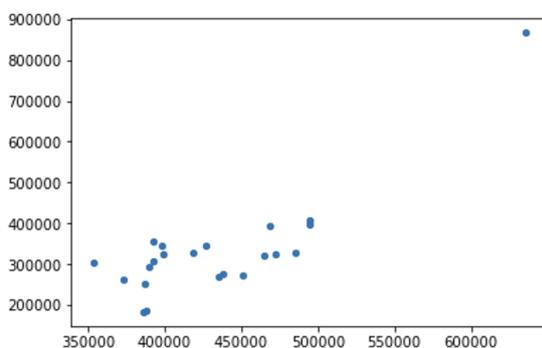
Hasil dan Pembahasan

Dataset hasil SUSENAS Maret 2020 yang digunakan yaitu rata-rata pengeluaran per kapita dalam sebulan menurut kelompok pengeluaran per kabupaten/kota. Data tersebut sudah merupakan pengelompokan berdasarkan kelompok pengeluaran yang merupakan kumulatif komoditas pengeluaran kabupaten/kota dan sudah merupakan hasil estimasi di setiap kabupaten/kota.

Penelitian ini menggunakan pendekatan data mining dalam prosesnya. Umumnya data harus melalui tahapan seperti *cleaning*, imputasi *missing values*, transformasi

data, *scaling* atau *normalizing*, dan lainnya. Namun dataset yang digunakan dalam penelitian sudah merupakan data yang siap digunakan dalam pengolahan dan merupakan rilis resmi BPS.

Sebagai analisis awal, dibuat *scatter plot* berdasarkan kolom total pengeluaran makanan dan kolom total pengeluaran bukan makanan dari seluruh kabupaten/kota dan didapatkan hasil seperti yang terlihat pada Gambar 3. Distribusi data menunjukkan sudah ada pengelompokan dari *dataset*. Data *outlier* (pencilan) dibagian kanan atas yang terpisah dari lainnya adalah data untuk Kota Kupang yang merupakan ibukota Provinsi Nusa Tenggara Timur.



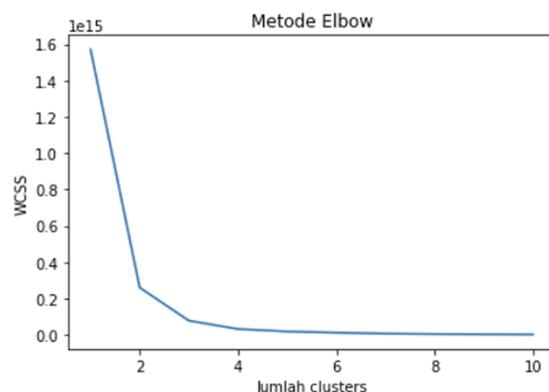
Gambar 3. Scatter Plot Total Pengeluaran Makanan dan Total Pengeluaran Non-Makanan.

Pengolahan Data

Sesuai dengan alur pembentukan *cluster*, langkah pertama adalah menentukan jumlah *cluster*. Metode yang digunakan pada penelitian ini adalah Elbow Methods. Muningsih (2017) menyatakan bahwa metode Elbow dapat digunakan untuk optimalisasi jumlah *cluster* pada K-

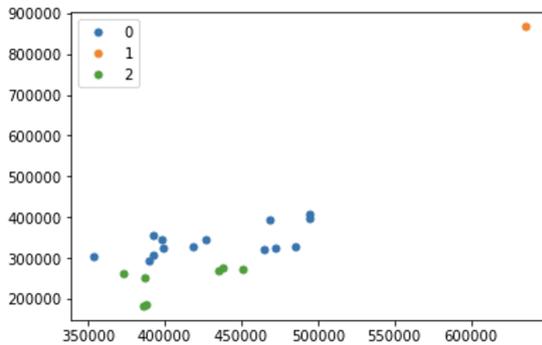
means dengan menghitung variasi tiap *cluster*. Metode Elbow ini menggunakan konsep yang disebut dengan *Within Cluster Sum of Square* (WCSS) yang merupakan nilai variasi dalam sebuah *cluster*. Hasil dari metode Elbow dataset ditunjukkan dalam Gambar 4. Dapat dilihat pada Gambar 4 bahwa jika ada 2 *cluster*, maka variasi dalam *cluster* akan cukup tinggi. Seiring dengan bertambah jumlah *cluster* maka WCSS akan makin kecil dan melandai setelah 4-5 *cluster*. Dapat disimpulkan bahwa berdasarkan dataset yang digunakan, jumlah *cluster* optimal adalah 4 atau 5 *cluster*.

Namun, tujuan penelitian ini ingin membagi habis seluruh wilayah kabupaten/kota di Provinsi NTT ke masing-masing kota Inflasi sehingga jumlah *cluster* ditetapkan berjumlah 3 yang mewakili masing-masing kota Inflasi. Sebagai informasi, berdasarkan SBH 2018 di Provinsi NTT ada 3 (tiga) Kota Inflasi yaitu Kota Kupang, Kota Maumere di Kabupaten Sikka, dan Kota Waingapu di Kabupaten Sumba Timur.



Gambar 4. Hasil Metode *Elbow* untuk Penentuan Jumlah *Cluster* Optimal

Setelah melakukan proses clustering dengan 3 *cluster*, hasilnya adalah sebagai berikut:



Gambar 5. Hasil Pengelompokan dengan 3 *Cluster*

Hasil pengelompokan per *cluster* sesuai wilayah kabupaten/kota di Provinsi NTT adalah sebagai berikut:

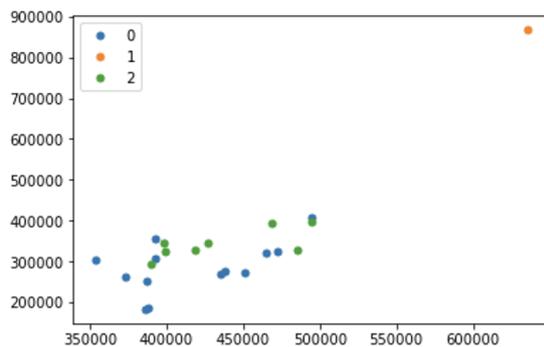
1. **Cluster pertama** terdiri atas Kabupaten Sumba Barat, Sumba Timur, Kupang, Belu, Lembata, Flores Timur, Sikka, Ende, Ngada, Manggarai, Rote Ndao, Manggarai Barat, Nagekeo, Sabu Raijua.
2. **Cluster kedua** terdiri dari Kota Kupang.
3. **Cluster Ketiga** terdiri dari Kabupaten Timor Tengah Selatan, Timor Tengah Utara, Alor, Sumba Tengah, Sumba Barat Daya, Manggarai Timur, Malaka.



Gambar 6. Peta Wilayah Provinsi NTT Berdasarkan Hasil *Cluster*.

Kota Kupang membentuk *cluster* sendiri karena keunikan pola konsumsi dan pengeluarannya. Hal ini wajar mengingat Kota Kupang merupakan ibu kota provinsi yang memiliki keragaman pola konsumsi dengan opsi komoditas yang lebih lengkap. Tetapi dalam iterasi ini, tidak ada kota Inflasi dalam *cluster* ketiga dan ada dua kota Inflasi masuk pada *cluster* pertama yaitu Maumere (Kabupaten Sikka) dan Waingapu (Kabupaten Sumba Timur).

Hasil ini belum memenuhi kriteria dimana dalam pemilihan *sister city* untuk penghitungan inflasi pada kota non-SBH, 7 wilayah pada *cluster* ketiga tidak akan bisa memilih kota Inflasi yang tepat. Dan sama halnya pada *cluster* pertama, 14 wilayah akan terbagi dalam dua pilihan kota inflasi dan tidak ada indikator penentuan untuk menjadi *sister city* salah satu kota Inflasi. Karena dataset dari Susenas Maret 2020 belum memenuhi tujuan penelitian, maka digunakan data tambahan yaitu data PDRB perkapita tahun 2020 seluruh kabupaten/kota di Provinsi NTT. Data ini ditambahkan karena menimbang data perekonomian makro yang memiliki konsep serupa dengan pengeluaran per kapita adalah data PDRB per kapita. Berikut hasilnya setelah ditambahkan dengan data PDRB per kapita:



Gambar 7. Hasil Pengelompokan dengan tiga *Cluster* dan Data Tambahan

Dalam iterasi ini, pengelompokan per *cluster* sesuai wilayah kabupaten/kota di Provinsi NTT adalah sebagai berikut:

1. ***Cluster* pertama** terdiri dari Kabupaten Timor Tengah Selatan, Timor Tengah Utara, Alor, Lembata, Sikka, Manggarai, Manggarai Barat, Sumba Tengah, Sumba Barat Daya, Nagekeo, Manggarai Timur, Sabu Raijua, Malaka.
2. ***Cluster* kedua** terdiri dari Kota Kupang.
3. ***Cluster* Ketiga** terdiri dari Kabupaten Sumba Barat, Sumba Timur, Kupang, Belu, Flores Timur, Ende, Ngada, Rote Ndao.



Gambar 8. Peta Wilayah Provinsi NTT Berdasarkan Hasil *Cluster*

Perubahan *cluster* terlihat setelah menambahkan data PDRB perkapita tahun 2020. Jumlah anggota pada *cluster* pertama berubah dari 14

wilayah menjadi 13 wilayah pada iterasi ini. Sedangkan anggota pada *cluster* ketiga berubah dari tujuh wilayah menjadi delapan wilayah.

Masing-masing Kota Inflasi kini sudah terpisah dalam *cluster* yang berbeda sehingga setiap Kabupaten/Kota sudah bisa menjadi *sister city* sesuai dengan kota Inflasi yang berada dalam *cluster* nya masing-masing. Kabupaten Timor Tengah Selatan, Timor Tengah Utara, Alor, Lembata, Manggarai, Manggarai Barat, Sumba Tengah, Sumba Barat Daya, Nagekeo, Manggarai Timur, Sabu Raijua, dan Malaka akan menggunakan Kota Maumere di Kabupaten Sikka sebagai acuan dalam pembentukan diagram timbang baru. Sedangkan Kabupaten Sumba Barat, Kupang, Belu, Flores Timur, Ende, Ngada, dan Rote Ndao dapat menggunakan Kota Waingapu di Kabupaten Sumba Timur sebagai acuan.

3. Simpulan dan Saran

Sister city adalah sebuah konsep yang digunakan oleh Kota Non-SBH untuk menghitung inflasi. Namun selama ini pemilihan *sister city* hanya berdasarkan kedekatan geografis dan asumsi bahwa pola konsumsi antara kota yang akan menghitung inflasi dan kota inflasi yang dipilih sebagai *sister city* memiliki kemiripan.

Saat ini, jika suatu daerah ingin menghitung IHK/inflasi di daerahnya sendiri, mereka akan memilih *sister city*

terdekat secara geografis. Untuk proyeksi anggaran, biasanya pemerintah daerah juga menggunakan angka inflasi dari *sister city* terdekat dalam formulasinya. Tetapi konsep *sister city* sendiri tidak didasari pembuktian secara ilmiah. Inilah yang mejadi dasar penelitian yaitu metode pemilihan *sister city* yang bisa dipertanggungjawabkan.

Memfaatkan machine learning yaitu teknik clustering dengan menggunakan algoritma K-Means, maka dilakukan penelitian untuk menentukan kelompok wilayah dengan wakil tiap kota Inflasi dalam clusternya. Dengan dataset Susenas Maret 2020, cluster telah terbentuk namun ada dua kota Inflasi yang terdapat dalam satu cluster sehingga tujuan penelitian belum tercapai. Oleh karena itu dilakukan penambahan data PDRB perkapita agar kota inflasi berada pada cluster yang berbeda.

Penelitian ini menghasilkan tiga cluster dimana masing-masing cluster mewakili kota Inflasi. Anggota dari tiap cluster nantinya bisa menggunakan kota Inflasi yang tergabung dalam

clusternya sebagai *sister city*. Dalam cluster Kota Kupang tidak memiliki anggota sehingga dapat disimpulkan bahwa kemungkinan tidak ada kemiripan kota-kota lainnya di NTT dengan kota Kupang sehingga tidak bisa dijadikan *sister city*. Hal ini kemungkinan terjadi karena sebagai ibu kota inflasi, Kota Kupang memiliki pola konsumsi yang unik dan berbeda dengan kabupaten lainnya dalam wilayah di NTT.

Pada tahun tahun 2022 akan dilakukan Survei Biaya Hidup 2022 yang merupakan dasar dalam penghitungan inflasi. Selain Kota Kupang, Maumere, dan Waingapu, ada dua kota SBH baru yaitu Soe di Kabupaten Timor Tengah Selatan dan Bajawa di Kabupaten Ngada. Hal ini menjadi opsi baru untuk pemilihan tentang *sister city* di masa depan. Selain itu, penggunaan lain machine learning dalam clustering seperti algoritma DB-SCAN dan juga modifikasi lebih lanjut seperti menggabungkan K-Means dan metode PCA (principal component analysis) juga dapat menjadi langkah baru dalam penelitian ke depannya.

Daftar Pustaka

- Alpaydin, E. (2020). *Introduction to Machine Learning, Fourth Edition*. <https://mitpress.mit.edu/books/introduction-machine-learning-fourth-edition>
- Badan Pusat Statistik. (2015). *Pedoman Penyusunan Diagram Timbang Dan Pengolahan Indeks Harga Konsumen Untuk Kabupaten/Kota Non Survei Biaya Hidup 2012*.
- Badan Pusat Statistik. (2017). *Penghitungan Inflasi Kabupaten*

- Lamongan Tahun 2017 Dengan Pendekatan Sister City*. <https://lamongankab.bps.go.id/news/2018/01/31/23/penghitungan-inflasi-kabupaten-lamongan-tahun-2017-dengan-pendekatan-sister-city--.html>
- Bank Indonesia. (n.d.). *Apa Itu Inflasi*. Retrieved October 29, 2021, from <https://www.bi.go.id/id/fungsi-utama/moneter/inflasi/default.aspx>
- Diacon, P.-E., & Maha, L.-G. (2015). The Relationship between Income, Consumption and GDP: A Time Series, Cross-Country Analysis. *Procedia Economics and Finance*, 23(October 2014), 1535–1543. [https://doi.org/10.1016/s2212-5671\(15\)00374-3](https://doi.org/10.1016/s2212-5671(15)00374-3)
- Hermawati, F. A. (2013). Data Mining. In *Data Mining*. Penerbit ANDI.
- Kholis, A. M., & Ferdian, A. (2019). Pengaruh Faktor-Faktor Ekonomi Terhadap Inflasi di Indonesia. *JIM UPB (Jurnal Ilmiah Manajemen Universitas Putera Batam)*, 7(1), 1. <https://doi.org/10.33884/jimupb.v7i1.900>
- Li, Y., & Wu, H. (2012). A Clustering Method Based on K-Means Algorithm. *Physics Procedia*, 25, 1104–1109. <https://doi.org/10.1016/j.phpro.2012.03.206>
- Mintz, Y., & Brodie, R. (2019). Introduction to artificial intelligence in medicine. *Minimally Invasive Therapy and Allied Technologies*, 28(2), 73–81. <https://doi.org/10.1080/13645706.2019.1575882>
- Muningsih, E. (2017). *Optimasi Jumlah Cluster K-Means Dengan Metode Elbow untuk Pemetaan Pelanggan* (pp. 105–114).
- Roihan, A., Sunarya, P., & Rafika, A. (2020). Pemanfaatan Machine Learning dalam Berbagai Bidang: Review paper. *IJCIT (Indonesian Journal on Computer and Information Technology)*, 5. <https://doi.org/10.31294/ijcit.v5i1.7951>
- Samuel, A. L. (1959). Some Studies in Machine Learning Using the Game of Checkers. *IBM Journal of Research and Development*, 3(3), 210–229. <https://doi.org/10.1147/rd.33.0210>
- Tim Pelaksana Kelompok Kerja Nasional TPID. (2014). *Buku Petunjuk TPID*. 39.
- Windarto, A. P. (2017). Penerapan Data Mining Pada Ekspor Buah-buahan Menurut Negara Tujuan Menggunakan K-Means Clustering. *Journal Teknologi Informasi: Techno.COM*, No. 4, 16(348–357).