

## BAB II. LANDASAN TEORI

### 2.1 Studi Literatur

Rusydi Umar, Abdul Fadlil, dan Rahmatika Az Zahra melakukan penelitian yang berjudul “*Self Organizing Maps (SOM)* untuk pengelompokan jurusan di SMK”. Penelitian yang dilakukan yaitu *clustering* dengan menggunakan algoritma SOM. *Clustering* merupakan pembelajaran tidak terbimbing atau *unsupervised learning* untuk melakukan pengelompokan. Pada penelitian ini, *clustering* diterapkan untuk mengelompokkan jurusan sekolah. Pengelompokan ini ditujukan kepada siswa yang bingung dalam memilih jurusan sekolah yang ingin dituju. Dalam penelitian ini, *clustering* dengan algoritma SOM dilakukan untuk membantu menyederhanakan penentuan/pengelompokan data calon siswa sesuai dengan parameter yang digunakan meliputi parameter keterampilan, bakat, dan minat calon siswa. Dalam penelitian ini, tidak terdapat hasil untuk uji akurasi. Namun, hasil dari pengelompokan dapat merekomendasikan jurusan yang dapat membantu calon siswa yang kebingungan karena merasa tidak memiliki bakat dan tidak pintar dalam beberapa materi pelajaran (Umar et al., 2018).

Penelitian yang dilakukan oleh Linda Maulida yang berjudul “Penerapan *Datamining* dalam Mengelompokkan Kunjungan Wisatawan Ke Objek Wisata Unggulan di Provinsi DKI Jakarta Dengan K-Means” didapatkan kesimpulan dengan menggunakan parameter jumlah pengunjung maka dapat dilakukan pengelompokan dan menghasilkan tiga *cluster* mengenai tingkat jumlah kunjungan wisatawan. Hasil *cluster* tersebut digunakan sebagai rekomendasi perbaikan sarana dan prasarana objek wisata unggulan, untuk meningkatkan jumlah kunjungan yang memiliki dampak pada promosi objek wisata serta peningkatan devisa negara (Maulida, 2018).

Pada penelitian yang dilakukan oleh Hartatik dan Arifa Satria Dwi Cahya pada tahun 2020 yang berjudul “Clusterisasi Kerusakan Gempa Bumi di Pulau Jawa Menggunakan SOM” menjelaskan mengenai penerapan metode SOM banyak dipakai untuk pengelompokan (*clustering*). Hal demikian karena metode SOM mendapatkan hasil yang dinilai stabil dilihat dari nilai *centroid* yang tidak mengalami perubahan pada setiap kali pengujian. (Hartatik & Cahya, 2020).

Dalam sebuah penelitian yang berjudul “Implementasi Self Organizing Maps (SOM) Klasifikasi Penduduk Untuk Menentukan Keputusan Pembangunan Daerah Prioritas Miskin (Studi Kasus Kota Makassar)” menyatakan bahwa algoritma SOM memiliki tingkat akurasi cukup baik untuk melakukan klasifikasi daerah, benda, tempat, geografis dan wilayah. Berdasarkan hasil analisis metode SOM baik digunakan dalam pembentukan *cluster* data dengan jumlah besar khususnya data statistik untuk menganalisa hasil dari visualisasi untuk mendapatkan informasi baru. Hasil visualisasi dari algoritma SOM juga dapat dianalisis untuk dijadikan bahan pertimbangan keputusan dari setiap kelas dan variabel yang dibentuk. Hasil akhir dari penelitian ini yaitu berupa sistem informasi klasifikasi tingkat kemiskinan yang menghasilkan informasi pemetaan penduduk miskin berdasarkan karakteristik demografi, geografi, pendidikan, ketenagakerjaan, perumahan, dan membantu pengambilan kebijakan dalam pemerataan program kemiskinan sehingga dapat melakukan pembangunan daerah secara prioritas (Muin, 2018). Oleh karena itu, pada penelitian ini metode SOM dipilih untuk melakukan pengelompokan terhadap data pariwisata karena di dalam data tersebut mengelompokkan objek wisata yang merupakan suatu geografis wilayah atau suatu tempat.

Penelitian yang berjudul “Implementasi Metode *K-Medoids Clustering* Untuk Mengetahui Pola Pemilihan Program Studi Mahasiswa Baru Tahun 2018 di Universitas Kanjuruhan Malang” menyatakan bahwa penelitian yang telah dilakukan mengenai pengelompokan mahasiswa baru dengan menggunakan metode *K-Medoids* dan menggunakan pengujian *silhouette coefficient* menghasilkan nilai yang yaitu sebesar 0.690754 dengan jumlah *cluster* sebanyak 3 dan jumlah data sebanyak 15 (Wira et al., 2019). Dari pernyataan diatas dapat disimpulkan bahwa teknik pengujian *silhouette coefficient* dapat menguji tingkat kualitas sebuah *cluster* dengan baik.

## **2.2 Dasar Teori**

### **2.2.1 Pariwisata**

Pariwisata adalah suatu kegiatan yang dilakukan seseorang dari tempat tinggalnya ke tempat lain, dengan waktu kunjungan yang berbeda. Pariwisata dapat mendorong pertumbuhan ekonomi yang pesat karena menyediakan lapangan kerja,

meningkatkan pendapatan, standar hidup, dan mendorong sektor produksi lainnya. Pariwisata dapat diartikan sebagai semua kegiatan wisata yang memiliki berbagai tujuan yang mengarah pada permintaan barang dan jasa selama perjalanan wisata. Segala kegiatan yang dilakukan oleh pemerintah untuk memenuhi dan menata kebutuhan wisatawan. Kegiatan tersebut berdampak pada perekonomian, sosial budaya, politik, pertahanan dan keamanan negara secara keseluruhan, sehingga kegiatan tersebut bermanfaat bagi kepentingan negara dan pembangunannya (Dr. M. Liga Suryadana., n.d.).

Di Indonesia, pariwisata menjadi salah satu sektor yang berperan penting dalam tergeraknya pembangunan perekonomian pada tingkat nasional maupun daerah. Hal tersebut dapat dilihat dari tingkat daya beli masyarakat serta kesejahteraan yang semakin tinggi membentuk pariwisata menjadi bagian penting atas kebutuhan maupun gaya hidup manusia. Pariwisata juga mengarahkan wisatawan guna mengenal alam dan budaya daerah lainnya (Bappeda, 2019).

Kabupaten probolinggo merupakan salah satu wilayah di Indonesia yang berlokasi di provinsi jawa timur. Karakteristik dan lokasinya yang memiliki keberagaman seperti dataran rendah, perbukitan, pegunungan, danau dan aliran sungai serta laut membuat Kabupaten Probolinggo memiliki potensi pariwisata yang cukup. Keberagaman karakteristik tersebut menciptakan tempat – tempat pariwisata yang mempunyai daya tarik pengunjung diantaranya wisata alam pegunungan, danau, wisata peninggalan sejarah, serta wisata seni dan budaya. Wisata–wisata tersebut menjadi sebuah modal yang potensial dalam pengembangan kepariwisataan di Kabupaten Probolinggo. Potensi tersebut perlu dikembangkan sehingga dapat menunjang pembangunan daerah serta pembangunan kepariwisataan khususnya (Bappeda, 2019).

### **2.2.2 Objek Wisata**

Objek wisata merupakan tempat yang dikunjungi orang untuk rekreasi/tempat berwisata. Objek wisata terdiri dari objek wisata alam, seperti pegunungan, danau, sungai, pantai, dan lautan, maupun berupa objek wisata bangunan, seperti museum, benteng, dan situs sejarah. Objek wisata merupakan bagian penting dari industri pariwisata. Di luar negeri objek wisata disebut sebagai *tourist attraction* (atraksi wisata), sedangkan di Indonesia disebut objek wisata (Laksana, 2020).

Kabupaten Probolinggo merupakan salah satu wilayah yang mengandalkan potensinya dalam hal pariwisata. Objek wisata yang ada di Kabupaten Probolinggo sangat beragam diantaranya yaitu Gunung bromo, Pantai bentar, Air terjun madakaripura, Arung jeram dan berbagai objek wisata lainnya yang sudah lama ada maupun yang masih baru berkembang (Badan Pusat Statistik (BPS), 2020). Beragamnya objek wisata yang ada di Kabupaten Probolinggo diharapkan dapat menarik minat pengunjung untuk datang menikmatinya.

Dalam tergeraknya objek wisata, pengunjung merupakan aspek yang paling penting. Pengunjung terbagi atas dua yaitu wisatawan nusantara yang merupakan pengunjung yang berasal dari dalam negeri dan wisatawan mancanegara yang merupakan pengunjung yang berasal dari luar negeri. Wisatawan yang berkunjung ke masing-masing objek wisata di Kabupaten Probolinggo bervariasi jumlahnya. Terdapat beberapa objek wisata yang mampu menyerap pengunjung hingga ratusan ribu orang per tahun. Adapun jumlah kunjungan wisatawan ke daerah wisata di Kabupaten Probolinggo, pada tahun 2017 Wisatawan Nusantara 655.867 orang dan wisatawan mancanegara yaitu sebesar 30.844 orang. Pada tahun sebelumnya mencapai 436.212 orang untuk wisatawan nusantara dan untuk wisatawan mancanegara mencapai 40.957 orang. Sehingga potensi pariwisata di Kabupaten Probolinggo sangat prospektif untuk terus dikembangkan, guna menyerap jumlah kunjungan wisatawan yang semakin meningkat dengan adanya pengembangan objek wisata (Bappeda, 2019).

Pengembangan objek wisata harus ditunjang dengan rencana pengembangan sarana dan prasarana wisata. Langkah-langkah yang harus diambil seperti melengkapi daya tarik wisata, terutama objek wisata prioritas dengan fasilitas penunjang wisata sesuai dengan karakter dari setiap objek wisata tersebut, dan karakter serta keinginan pengunjung ketika mendatangi lokasi (Bappeda, 2019).

### **2.2.3 Clustering**

*Clustering* adalah pembelajaran tanpa pengawasan atau *unsupervised learning* (Umar et al., 2018). Dengan kata lain *clustering* bertujuan untuk mengelompokkan objek-objek ke dalam *cluster-cluster*. *Cluster* adalah sekelompok atau sekumpulan objek data yang sejenis satu sama lain. Objek-objek akan dikelompokkan ke dalam satu atau lebih *cluster* sehingga objek-objek yang berada

dalam suatu klaster akan mempunyai kesamaan yang tinggi antara satu dengan yang lainnya.

Winarto menjelaskan bahwa analisis pengelompokan (*clustering*) merupakan proses membagi data dalam suatu himpunan menjadi beberapa kelompok yang memiliki kesamaan data yang berbeda antara satu kelompok dengan kelompok yang lainnya (Maulida, 2018). Potensi *clustering* dapat digunakan untuk mengetahui struktur dalam data yang dapat dipakai lebih lanjut dalam berbagai aplikasi secara luas seperti klasifikasi, pengolahan gambar, dan pengenalan pola (Maulida, 2018).

Pada proses analisis *cluster* metode yang digunakan untuk membagi data menjadi *subset* data berdasarkan kemiripan yang telah ditentukan sebelumnya. Jadi analisis *cluster* secara umum dapat dikatakan bahwa:

- a. Data yang terdapat dalam satu cluster memiliki tingkat kesamaan yang tinggi
  - b. Data dalam suatu cluster yang berbeda memiliki tingkat kesamaan yang rendah
- (Maulida, 2018).

#### **2.2.4 Metode Self Organizing Maps (SOM)**

*Self Organizing Map* (SOM) disebut juga dengan *neural network kohonen* yang efektif digunakan sebagai analisis data multidimensi. Untuk melakukan analisis *cluster* dengan mempertahankan data struktur (topologi) dapat menggunakan jaringan SOM, karena pada lapisan keluaran jaringan masukan data tetap dekat dan bersama-sama (Muin, 2018).

Jaringan *kohonen* SOM diperkenalkan oleh Teuvo Kohonen seorang ilmuwan Finlandia pada tahun 1982. SOM merupakan model dari pendekatan jaringan syaraf tiruan. Jaringan kohonen merupakan salah satu bentuk topologi dari *Unsupervised Artificial Neural Network* (*Unsupervised ANN*) hal tersebut karena dalam proses pemeliharaan tidak memerlukan pengawasan (*target output*). SOM mengenali penyebaran (seperti *competitive learning*) dan topologi dari vektor masukan yang melewati proses *training*. Larose menjelaskan bahwa SOM memiliki tiga karakteristik meliputi kompetisi yaitu setiap vektor bobot saling berlomba untuk menjadi simpul pemenang, kemudian kooperasi yaitu kerja sama antara setiap simpul pemenang dengan lingkungannya, dan adaptasi yaitu perubahan yang terjadi oleh simpul pemenang (Kashi et al., 2018).

Algoritma pembelajaran *unsupervised* pada *Kohonen SOM* dalam pengelompokan data adalah sebagai berikut (Umar et al., 2018):

1. Menentukan *learning rate*.
2. Inisialisasi bobot ( $w_{ij}$ ) dengan nilai random antara 0 sampai 1,
3. Melakukan input data.
4. Menghitung jarak dengan persamaan *Euclidean Distance* melalui persamaan (2.1)

$$d(j) = \sqrt{\sum_{j=0}^{i=1} (w_{ij} - x_i(t))^2} \quad (2.1)$$

Dimana :

$d$  = jarak

$x_i(t)$  = node data input

$w_{ij}$  = bobot ke- $ij$

5. Mencari nilai terkecil dari seluruh bobot ( $d_j$ ). Index dari bobot ( $d_j$ ) yang paling mirip disebut *winning neuron*.
6. Memperbarui setiap bobot  $ij$  dengan menggunakan persamaan (2.2):

$$w_{ij}(t + 1) = w_{ij}(t) + \alpha(t) \cdot [x_i(t) - w_{ij}(t)] \quad (2.2)$$

Dimana :

$x_i$  = data input

$w_{ij}$  = bobot

$t$  = waktu

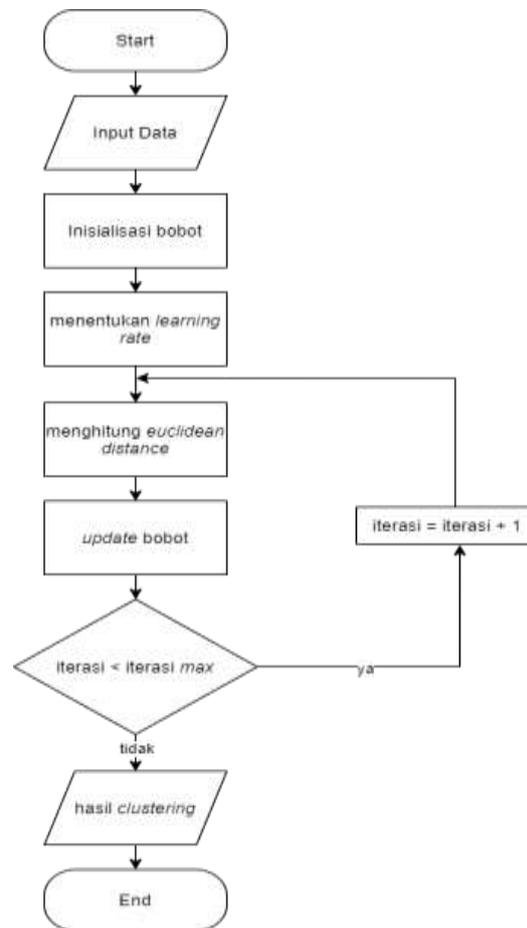
$i$  = index inputan

$j$  = index output

$\alpha$  = *alpha learning rate*

9. Simpan bobot hasil.

Gambar dibawah ini merupakan langkah-langkah perhitungan SOM dalam bentuk diagram *flowchart*.

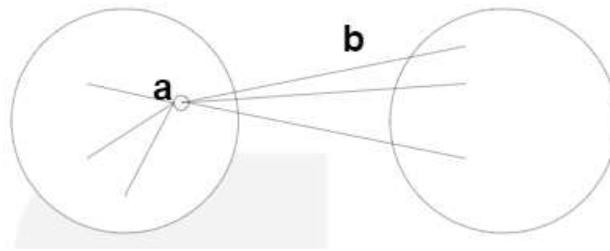


Gambar 2. 1 *Flowchart* SOM

### 2.2.5 *Silhouette Coefficient*

*Silhouette Coefficient* merupakan metode pengujian untuk validasi *cluster* terhadap objek yang ada di dalamnya. Metode ini digunakan untuk menguji kualitas dari suatu *cluster* yang dihasilkan dari sebuah metode *clustering*. Metode *Silhouette coefficient* menggabungkan dua metode yaitu metode *cohesion* untuk mengukur kedekatan relasi antar objek dalam suatu *cluster* dan metode *separation* untuk mengukur seberapa jauh *cluster* terpisah dengan *cluster* lain (Pramesti et al., 2017).

Berikut adalah pemaparan dalam menentukan *silhouette coefficient* dari sebuah *cluster*:



Gambar 2. 2 Ilustrasi *Silhouette Coefficient* (sumber : (Putra et al., 2015))

Pada pada gambar 2.1 diilustrasikan terdapat sebuah titik  $x$  yang berada pada *cluster* A, lalu  $a(i)$  merupakan jarak rata – rata antara titik  $x$  engan titik lain yang terdapat pada *cluster* A, dan  $b(i)$  merupakan jarak rata – rata antara titik  $x$  dengan titik lain di *cluster* lain yang terdekat dengan *cluster* A, yaitu *cluster* B.

Berikut adalah tahapan perhitungan *silhouette coefficient* (Handoyo et al., 2014) :

1. Menghitung rata – rata jarak dari objek  $i$  dengan seluruh objek yang berada dalam satu *cluster*, maka didapatkanlah nilai rata – rata yang diset  $a(i)$ . Rumus yang digunakan untuk mencari jarak antara objek  $i$  dengan objek lain yaitu dengan rumus *euclidean distance*.

$$a(i) = \frac{1}{|A|-1} \sum_{j \in A, j \neq i} d(i, j) \quad (2.3)$$

2. Menghitung rata – rata jarak objek  $i$  dengan objek yang berada pada *cluster* lain, kemudian diambil nilai terkecilnya.

$$d(i, C) = \frac{1}{|C|} \sum_{j \in C} d(i, j) \quad (2.4)$$

3. Untuk menghitung nilai *Silhouette Coefficient*nya digunakan rumus sebagai berikut :

$$S(i) = \frac{b(i)-a(i)}{\max(a(i), b(i))} \quad (2.5)$$

Hasil perhitungan nilai *Silhouette Coefficient* bervariasi antara -1 sampai 1. Hasil *clustering* dikatakan baik bila nilai *Silhouette Coefficient* berupa nilai yang positif. Apabila nilai yang dihasilkan 1 maka objek tersebut sudah berada di *cluster* yang tepat. Apabila nilai yang dihasilkan berupa angka 0 maka objek tersebut tidak jelas harus masuk *cluster* A atau *cluster* B. Namun, apabila nilai yang dihasilkan berupa -1 maka objek tersebut *overlapping* yang berarti tidak cocok pada *cluster* tersebut.