

BAB II

LANDASAN TEORI

2.1 Beasiswa

Menurut Kamus Besar Bahasa Indonesia Beasiswa adalah tunjangan yang diberikan kepada pelajar atau mahasiswa sebagai bantuan biaya belajar. Pemberian beasiswa kepada mahasiswa merupakan program kerja yang ada di setiap universitas atau perguruan tinggi. Program beasiswa diadakan untuk meringankan beban mahasiswa dalam biaya kuliah. Pemberian beasiswa kepada mahasiswa dilakukan secara selektif sesuai dengan jenis beasiswa yang diadakan.

Menurut (Dewi, Narayana and Wibawa, 2019), beasiswa diartikan sebagai bentuk penghargaan yang diberikan kepada individu agar dapat melanjutkan pendidikan ke jenjang yang lebih tinggi. Penghargaan itu dapat berupa akses tertentu pada suatu institusi atau penghargaan berupa bantuan keuangan. Pada dasarnya, beasiswa adalah penghasilan bagi yang menerimanya. Hal ini sesuai dengan ketentuan pasal 4 ayat (1) Undang-undang PPh/2000. Disebutkan pengertian penghasilan adalah tambahan kemampuan ekonomis dengan nama dan dalam bentuk apa pun yang diterima atau diperoleh dari sumber Indonesia atau luar Indonesia yang dapat digunakan untuk konsumsi atau menambah kekayaan Wajib Pajak. Karena beasiswa bisa diartikan menambah kemampuan ekonomis bagi penerimanya, berarti beasiswa merupakan penghasilan.

Beasiswa dapat diberikan oleh lembaga pemerintah, perusahaan ataupun yayasan. Pemberian beasiswa dapat dikategorikan pada pemberian cuma-cuma ataupun pemberian dengan ikatan kerja (biasa disebut ikatan dinas) setelah selesainya pendidikan. Lama ikatan dinas ini berbeda-beda, tergantung pada lembaga yang memberikan beasiswa tersebut. beasiswa juga banyak diberikan kepada per kelompok (*group*) misalnya ketika ada *event* perlombaan yang diadakan oleh lembaga pendidikan, dan salah satu hadiahnya adalah beasiswa.

2.1.1 Jenis-Jenis Beasiswa

Menurut (Qoiriyah, Purwanto and Setyaningsih, 2019), ada beberapa jenis beasiswa yaitu :

a. Beasiswa Penghargaan

Beasiswa ini biasanya diberikan kepada kandidat yang memiliki keunggulan akademik. Beasiswa ini diberikan berdasarkan prestasi akademik mereka secara keseluruhan. Misalnya, dalam bentuk Indeks Prestasi Kumulatif (IPK). Meski sangat kompetitif, beasiswa ini ada dalam berbagai bentuk.

b. Beasiswa Bantuan

Jenis beasiswa ini adalah untuk mendanai kegiatan akademik para mahasiswa yang kurang beruntung, tetapi memiliki prestasi. Komite beasiswa biasanya memberikan beberapa penilaian pada kesulitan ini, misalnya, seperti pendapatan orangtua, jumlah saudara kandung yang sama-sama tengah menempuh studi, pengeluaran, biaya hidup, dan lain-lain.

c. Beasiswa Atletik

Universitas biasanya merekrut atlet populer untuk diberikan beasiswa dan dijadikan tim atletik perguruan tinggi mereka. Banyak atlet menyelesaikan pendidikan mereka secara gratis, tetapi membayarnya dengan prestasi olahraga. Beasiswa seperti ini biasanya tidak perlu dikejar, karena akan diberikan kepada mereka yang memiliki prestasi.

d. Beasiswa Penuh

Banyak orang menilai bahwa beasiswa diberikan kepada penerimanya untuk menutupi keperluan akademik secara keseluruhan. Jika Anda benar-benar beruntung, tentunya Anda akan mendapatkan beasiswa seperti ini. Beasiswa akan diberikan untuk menutupi kebutuhan hidup, buku, dan biaya pendidikan. Namun, banyak beasiswa lainnya meng-cover biaya hidup, buku, atau sebagian dari uang sekolah.

2.1.2 Tujuan Beasiswa

Menurut Dikti dalam pedoman umum beasiswa Bantuan Biaya Akademik (BBA) dan Peningkatan Prestasi Akademik (PPA) tahun 2019. Tujuan diselenggarakannya beasiswa antara lain :

- a. Meningkatkan prestasi mahasiswa penerima baik kurikuler, maupun ekstrakurikuler serta motivasi berprestasi bagi mahasiswa lain.
- b. Mengurangi jumlah mahasiswa yang putus kuliah, karena tidak mampu membiayai pendidikan.
- c. Meningkatkan akses dan pemerataan kesempatan belajar di perguruan tinggi.

2.1.3 Prioritas Penentuan Beasiswa

Prioritas kriteria yang digunakan untuk penentuan mahasiswa penerima beasiswa tidak sama untuk setiap jenis beasiswa. Sebagai contoh urutan prioritas kriteria penerima beasiswa PPA dan BBA adalah :

a. Prioritas Penerima PPA

Apabila calon penerima melebihi kuota yang telah ditetapkan, maka perguruan tinggi dapat menentukan mahasiswa penerima sesuai dengan urutan prioritas sebagai berikut :

1. Mahasiswa yang memiliki IPK paling tinggi.
2. Mahasiswa yang memiliki SKS paling banyak dalam satu angkatan.
3. Mahasiswa yang memiliki prestasi pada kegiatan ko/ekstra kurikuler (penalaran, minat, dan bakat) tingkat internasional/dunia, Regional/Asia/Asean, dan Nasional
4. Mahasiswa yang memiliki keterbatasan kemampuan ekonomi.

b. Prioritas Penerima BBA

Apabila calon penerima melebihi kuota yang telah ditetapkan, maka perguruan tinggi dapat menentukan mahasiswa penerima sesuai dengan urutan prioritas sebagai berikut :

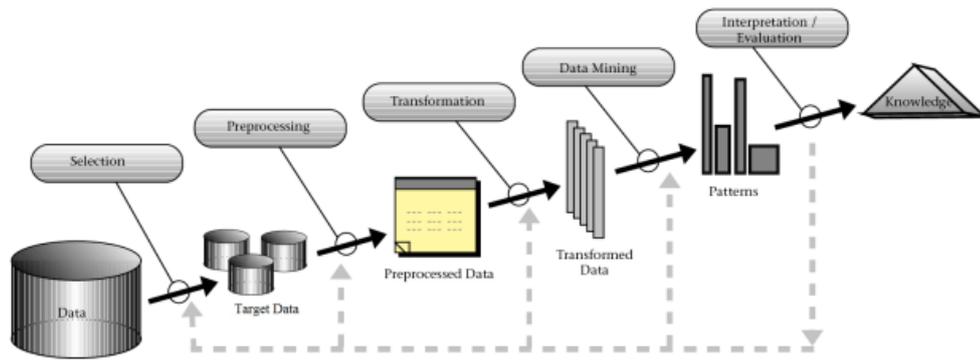
1. Mahasiswa dengan orangtua/penanggung biaya kuliah mengalami kendala finansial, dari keluarga peserta Program Keluarga Harapan (PKH), atau keluarga pemegang Kartu Keluarga Sejahtera (KKS)
2. Mahasiswa yang tidak sedang dibiayai oleh Program Bidikmisi atau program beasiswa lainnya
3. Mahasiswa yang sedang menjalani perkuliahan di semester 3, 5, dan 7.

2.2 Data Mining

Data mining semakin populer dan berkembang dengan tren big data seiring dengan mudahnya mengakses informasi yang semakin luas dari waktu ke waktu. Data mining merupakan proses pencarian, pola, korelasi, dan tren yang menggabungkan bidang machine learning, teknik visualisasi, dan statistika melalui penyaringan dari sebuah data yang besar. Data mining juga menjadi berkembang dengan seiringnya tren baru dalam dunia teknologi informasi yang bertujuan untuk mendefinisikan data-data penting dan pengetahuan yang berada dalam sistem informasi (Toro, R., & Lestari, S. 2023).

Istilah data mining sering digunakan sebagai sinonim untuk proses menemukan pola-pola yang berwawasan, menarik, dan baru, serta model deskriptif, dapat dipahami, dan prediktif dari data skala besar. Menurut (Wanto *et al.*, 2020) mendefinisikan data mining sebagai proses untuk mendapatkan informasi yang berguna dari gudang basis data yang besar, yang dapat juga diartikan sebagai pengekstrakan informasi baru yang diambil dari bongkahan data besar yang membantu dalam pengambilan keputusan (*Knowledge Discovery*). Pendekatan dasar dalam data mining adalah untuk meringkas data dan untuk mengekstrak informasi berguna yang masuk akal dan sebelumnya tidak diketahui.

Menurut (Arhami and Nasir, 2020), data mining merupakan proses penggalian informasi dan pola yang bermanfaat dari data yang sangat besar. Data mining mencakup pengumpulan data, ekstraksi data, analisis data, dan statistik data. Data mining juga dikenal sebagai *knowledge discovery*, *knowledge extraction*, *data/pattern analysis*, *information harvesting*, dan lain-lain. Proses KDD (*Knowledge Discovery in Databases*) secara garis besar dapat dijelaskan pada gambar 2.1.



Gambar 2.1 Tahapan Data Mining (Wanto et al., 2020)

a. Data Selection

Pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam *knowledge data discovery* (KDD) dimulai. Data hasil seleksi yang akan digunakan untuk proses data mining, disimpan dalam suatu berkas, terpisah dari basis data operasional.

b. *Preprocessing* atau *Cleaning*

Sebelum proses data mining dapat dilaksanakan, perlu dilakukan proses *cleaning* pada data yang menjadi fokus *knowledge data discovery*. Proses *cleaning* mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak juga dilakukan proses enrichment, yaitu proses memperkaya data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan untuk KDD, seperti data atau informasi.

c. *Transformation*

Coding adalah proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses data mining. Proses *coding* dalam *knowledge data discovery* merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

d. Data mining

Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam data mining sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat tergantung pada tujuan dan proses KDD secara keseluruhan.

e. *Interpretation* atau *Evaluation*

Pola informasi yang dihasilkan dari proses data mining perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang disebut *interpretation*. Tahap ini mencakup pemeriksaan apakah pola informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada pada sebelumnya

Ada beberapa tugas yang dapat dilakukan oleh data mining dalam proses pemecahan masalah dan pencarian pengetahuan baru, diantaranya adalah sebagai berikut:

a. *Klustering (Clustering)*

Digunakan untuk mengelompokan atau mengidentifikasi data yang memiliki karakteristik tertentu. Contoh algoritma : *K-Means*, *K-Medoids*, dan lain-lain.

b. *Klasifikasi (Classification)*

Digunakan untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui. Contoh algoritma: *C4.5*, *K-Nearest Neighbor (KNN)*, dan lain-lain.

c. *Asosiasi (Association)*

Digunakan untuk mengatasi masalah bisnis yang khas, yakni dengan menganalisa tabel transaksi penjualan dan mengidentifikasi produk-produk yang seringkali dibeli bersamaan oleh customer, misalnya apabila orang membeli sambal, biasanya juga dia membeli kecap. Contoh Algoritma: *Apriori*, *Frequent Pattern Growth (FP- Growth)*, dan lain-lain.

d. *Estimasi (Estimation)*

Digunakan untuk memperkirakan atau menilai sesuatu hal yang belum pernah ada sebelumnya yang disajikan dalam bentuk hasil kuantitatif (angka). Contoh algoritma: *Regresi Linier*, *Confidence Interval Estimations*, dan lain-lain.

e. *Prediksi (Predictions)*

Digunakan untuk memperkirakan atau meramalkan suatu kejadian yang belum pernah terjadi. Contoh algoritma : *Decision Tree*, *K-Nearest Neighbor (KNN)*, dan lain-lain.

2.3 Clustering

Clustering adalah suatu pekerjaan mengelompokkan sekumpulan objek data sehingga objek – objek dalam suatu kelompok memiliki kemiripan yang tinggi, tetapi sangat berbeda dengan objek – objek dalam kelompok lain. Proses pada *Clustering* akan mempartisi sekumpulan objek data menjadi subset yang dapat di manfaatkan untuk mengatur hasil pencarian dalam kelompok – kelompok dan menyajikan hasil secara ringkas dan mudah di akses (Nisar, N., & Kusumajaya, H. 2022).

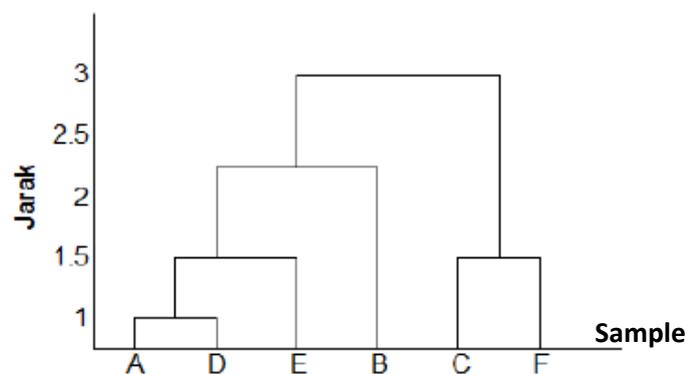
Menurut (Rahmayani, 2018), *clustering* adalah proses pengelompokkan kumpulan data menjadi beberapa kelompok sehingga objek di dalam satu kelompok memiliki banyak kesamaan dan memiliki banyak perbedaan dengan objek dikelompok lain. Perbedaan dan persamaannya biasanya berdasarkan nilai atribut dari objek tersebut dan dapat juga berupa perhitungan jarak. *Clustering* sendiri juga disebut *Unsupervised Classification*, karena *clustering* lebih bersifat untuk dipelajari dan diperhatikan. *Cluster analysis* merupakan proses partisi satu set objek data ke dalam himpunan bagian. Setiap himpunan bagian adalah cluster, sehingga objek yang di dalam cluster mirip satu sama dengan yang lainnya, dan mempunyai perbedaan dengan objek dari cluster yang lain. Partisi tidak dilakukan dengan manual tetapi dengan algoritma *clustering*. Oleh karena itu, *Clustering* sangat berguna dan bisa menemukan group yang tidak dikenal dalam data.

Dengan menggunakan *clustering*, kita dapat mengidentifikasi daerah yang padat, menemukan pola-pola distribusi secara keseluruhan, dan menemukan keterkaitan yang menarik antara atribut-atribut data. Dalam data mining, usaha difokuskan pada metode-metode penemuan untuk *cluster* pada basis data berukuran besar secara efektif dan efisien. Beberapa kebutuhan *clustering* dalam data mining meliputi skalabilitas, kemampuan untuk menangani tipe atribut yang berbeda, mampu menangani dimensionalitas yang tinggi, menangani data yang mempunyai

noise, dan dapat diterjemahkan dengan mudah. *Clustering* diklasifikasikan menjadi dua metode, yaitu *hierarchy clustering* dan *non hierarchy clustering*.

2.3.1 Metode Hierarchy Clustering

Metode *hierarchy clustering*, ialah metode yang memulai pengelompokannya dengan dua atau lebih objek yang mempunyai kesamaan paling dekat, kemudian proses dilanjutkan ke objek lain yang mempunyai kedekatan kedua. Demikian seterusnya sehingga cluster akan membentuk seperti “pohon” dimana ada *hierarchy* (tingkatan) yang jelas antar objek, dari yang paling mirip sampai dengan yang paling tidak mirip. Secara logika semua objek pada akhirnya hanya akan membentuk sebuah cluster. Dendogram *clustering* biasanya digunakan untuk membantu memperjelas proses hierarki tersebut. Contoh dendogram *clustering* dapat dilihat pada gambar 2.2. Metode *hierarchy clustering* dibagi lagi menjadi beberapa metode yaitu Pautan Tunggal (*Single Linkage*), Pautan Lengkap (*Complete Linkage*), Pautan Rata-rata (*Average Linkage*), Metode Ward (*Ward's Method*) dan Metode Centroid (*pusat*).



Gambar 2.2 Dendogram (Sembiring, Zain and Embong, 2011)

2.3.2 Metode Non Hierarchy Clustering

Metode *Non Hierarchy Clustering* adalah metode yang dimulai dengan menentukan terlebih dahulu jumlah cluster yang diinginkan (dua *cluster*, tiga *cluster* atau yang lain) dan kemudian baru dilakukan proses cluster tanpa mengikuti proses hierarki yang biasa disebut metode *K-Means Cluster*. Dua kelemahan dari prosedur non-hierarki ialah bahwa banyaknya *cluster* harus

disebutkan/ditentukan sebelumnya dan pemilihan pusat *cluster* sembarang. Hasil *cluster* tergantung pada bagaimana pusat dipilih. Banyak program non-hierarki, memilih k objek (kasus) yang pertama, tanpa ada nilai yang hilang sebagai pusat *cluster* awal (k = banyaknya *cluster*). Jadi hasil *cluster* tergantung pada urutan observasi dalam data. Metode *Non Hierarchy Clustering* biasanya dipilih untuk melakukan pengelompokan objek dengan sampel data yang besar.

2.3.3 Syarat Clustering

Menurut (Irwansyah and Faisal, 2015) syarat sekaligus tantangan yang harus dipenuhi oleh suatu algoritma *clustering* adalah :

a. Skalabilitas

Suatu metode *clustering* harus mampu menangani data dalam jumlah yang besar. Saat ini data dalam jumlah besar sudah sangat umum digunakan dalam berbagai bidang misalnya saja suatu database. Tidak hanya berisi ratusan objek, suatu *database* dengan ukuran besar bahkan berisi lebih dari jutaan objek.

b. Kemampuan analisa beragam bentuk data

Algoritma klasteriasi harus mampu diimplementasikan pada berbagai macam bentuk data seperti data nominal, ordinal maupun gabungannya.

c. Menemukan *cluster* dengan bentuk yang tidak terduga

Banyak algoritma *clustering* yang menggunakan metode *Euclidean* atau *Manhattan* yang hasilnya berbentuk bulat. Padahal hasil *clustering* dapat berbentuk aneh dan tidak sama antara satu dengan yang lain. Karenanya dibutuhkan kemampuan untuk menganalisa cluster dengan bentuk apapun pada suatu algoritma *clustering*.

d. Kemampuan untuk dapat menangani noise

Data tidak selalu dalam keadaan baik. Ada kalanya terdapat data yang rusak, tidak dimengerti atau hilang. Karena sistem inilah suatu algoritma *clustering* dituntut untuk mampu menangani data yang rusak.

e. Sensitifitas terhadap perubahan *input*

Perubahan atau penambahan data pada *input* dapat menyebabkan terjadi perubahan pada *cluster* yang telah ada bahkan bisa menyebabkan perubahan yang mencolok apabila menggunakan algoritma clustering yang memiliki tingkat sensitifitas rendah.

f. Mampu melakukan *clustering* untuk data dimensi tinggi

Suatu kelompok data dapat berisi banyak dimensi ataupun atribut. Untuk itu diperlukan algoritma *clustering* yang mampu menangani data dengan dimensi yang jumlahnya tidak sedikit.

g. Interpretasi dan kegunaan

Hasil dari *clustering* harus dapat diinterpretasikan dan berguna.

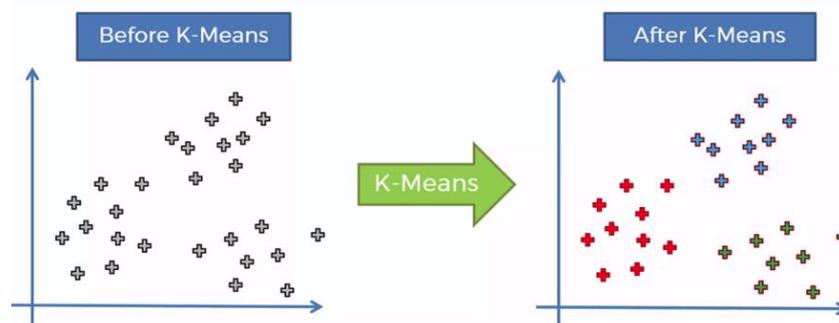
2.4 Algoritma *K-Means*

K-Means Clustering merupakan metode analisis cluster yang bertujuan untuk memecah objek menjadi K Cluster kemudian di amati di mana setiap objek cluster di peroleh melalui rata-rata terdekat. Algoritma ini merupakan salah satu pembelajaran yang terkenal sederhana dan mudah dipelajari sebagai pemecah masalah pengelompokkan dari sebuah dataset. Algoritma *K-Means* adalah algoritma evolusi yang mana metode operasinya memiliki artian yang serupa dengan nama algoritmanya (Dedi A, Sriyanto., & Joko T., 2023)

K-Means merupakan salah satu algoritma *clustering* yang masuk dalam kelompok *unsupervised learning* yang digunakan untuk membagi data menjadi beberapa kelompok dengan sistem partisi. Algoritma ini menerima masukan berupa data tanpa label kelas. Pada algoritma *K-Means*, komputer mengelompokkan sendiri data-data yang menjadi masukannya tanpa mengetahui terlebih dahulu target kelasnya. Masukan yang diterima adalah data atau objek dan k buah kelompok (*cluster*) yang diinginkan. Algoritma ini akan mengelompokkan data atau objek kedalam k buah kelompok tersebut.

Pada setiap *cluster* terdapat titik pusat (*centroid*) yang mempresentasikan *cluster* tersebut. Pada pemrosesan data algoritma *K-Means clustering*, akan diawali dengan pengelompokan *centroid* pertama yang dipilih secara acak sebagai titik awal untuk setiap *cluster*, kemudian menghitung secara berulang agar posisi *centroid* optimal.

Pada dasarnya algoritma K-Means hanya mengambil sebagian dari banyaknya komponen yang didapatkan untuk kemudian dijadikan pusat *cluster* awal, pada penentuan pusat *cluster* ini dipilih secara acak dari populasi data. Kemudian algoritma *K-Means* akan menguji masing-masing dari setiap komponen dalam populasi data tersebut dan menandai komponen tersebut ke dalam salah satu pusat *cluster* yang telah didefinisikan sebelumnya tergantung dari jarak minimum antar komponen dengan tiap-tiap pusat *cluster*. Selanjutnya posisi pusat *cluster* akan dihitung kembali hingga semua komponen data digolongkan ke dalam tiap-tiap *cluster* dan terakhir akan terbentuk *cluster* baru (Wanto *et al.*, 2020).



Gambar 2.3 K-Means Clustering (Wanto *et al.*, 2020)

K-Means merupakan salah satu metode data *clustering non-hierarchical* atau *partitional clustering*. Algoritma *K-Means* berusaha mengelompokkan data yang ada ke dalam beberapa kelompok, di mana data dalam satu kelompok mempunyai karakteristik yang sama antara satu dengan yang lainnya dan mempunyai karakteristik yang berbeda dengan data yang ada di dalam kelompok yang lain. Langkah-langkah algoritma K-Means dapat dijelaskan sebagai berikut :

- a. Tentukan jumlah *cluster* (k) pada data set
- b. Tentukan nilai pusat (*centroid*)

Penentuan nilai *centroid* pada tahap awal dilakukan secara random, sedangkan pada tahap iterasi digunakan rumus seperti pada persamaan (1) berikut ini.

$$V_{ij} = \frac{1}{N_i} \sum_{K=0}^{N_i} X_{kj} \dots\dots\dots(1)$$

Keterangan :

- V_{ij} = *centroid* rata-rata *cluster* ke-I untuk variabel ke-j
 N_i = jumlah anggota *cluster* ke-i
 i, k = indeks dari *cluster*
 j = indeks dari variable
 X_{kj} = nilai data ke-k variabel ke-j untuk *cluster* tersebut

- c. Pada masing-masing *record*, hitung jarak terdekat dengan *centroid*. Ada beberapa cara yang dapat digunakan untuk mengukur jarak data ke pusat kelompok, diantaranya *euclidean*, *manhattan/city block*, dan *minkowsky*. Setiap cara memiliki kelebihan dan kekurangan masing-masing. Pada penulisan ini, jarak *centroid* yang digunakan adalah *euclidean distance*. Perhitungan jarak menggunakan Persamaan (2)

$$De = \sqrt{(xi - si)^2 + yi - ti^2} \dots\dots\dots(2)$$

Keterangan :

- De = *euclidean distance*
 i = banyaknya objek ²
 (x, y) = koordinat objek
 (s, t) = koordinat *centroid*

- d. Kelompokkan objek berdasarkan jarak ke *centroid* terdekat
 e. Ulangi langkah c hingga langkah d , lakukan iterasi hingga *centroid* bernilai optimal.

2.5 Jaringan Syaraf Tiruan

Jaringan Syaraf Tiruan merupakan model yang meniru cara kerja jaringan biologis. Salah satu contoh pengambilan ide dari jaringan syaraf biologis adalah adanya elemen-elemen pemrosesan pada jaringan syaraf tiruan yang saling terhubung dan beroperasi secara paralel dan meniru jaringan saraf biologis yang tersusun dari ses-sel syaraf neuron (Suparji, S., Sriyanto, S., & Lestari, S. 2022).

Jaringan Syaraf Tiruan tercipta sebagai suatu generalisasi model matematis dari pemahaman manusia (*human cognition*) yang didasarkan atas asumsi sebagai berikut:

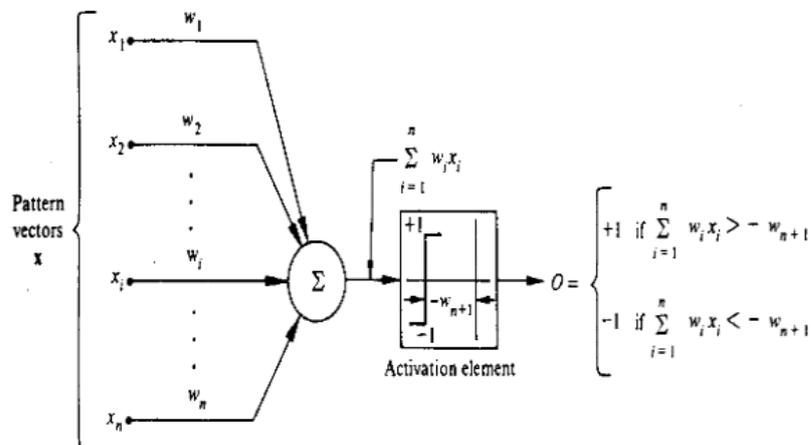
- a. Pemrosesan informasi terjadi pada elemen sederhana yang disebut neuron.

- b. Isyarat mengalir diantara sel syaraf/neuron melalui suatu sambungan penghubung.
- c. Setiap sambungan penghubung memiliki bobot yang bersesuaian.
- d. Setiap sel syaraf akan merupakan fungsi aktivasi terhadap isyarat hasil penjumlahan berbobot yang masuk kepadanya untuk menentukan isyarat keluarannya.

Menurut (Nafi'iyah, 2015) jaringan syaraf terdiri atas beberapa neuron dan ada hubungan antara neuron-neuron tersebut. Neuron adalah sebuah unit pemroses informasi yang menjadi dasar pengoperasian jaringan syaraf tiruan. Syaraf adalah sebuah unit pemroses informasi dengan tiga elemen dasar yaitu :

- a. Satu set link yang terhubung.
- b. Sebuah penjumlah untuk menghitung besarnya penambahan pada sinyal masukan.
- c. Sebuah fungsi aktivasi untuk membatasi banyaknya keluaran pada syaraf.

Model jaringan syaraf tiruan dapat dilihat pada Gambar 2.4.



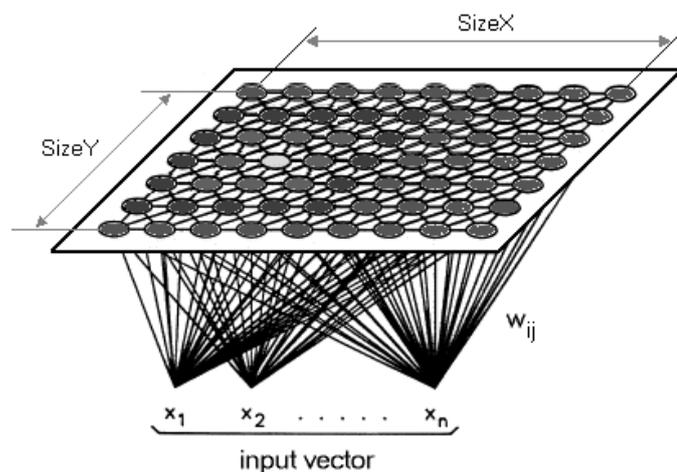
Gambar 2.4 Model Jaringan Syaraf Tiruan (Nafi'iyah, 2015)

Setiap pola-pola informasi input dan output yang diberikan ke dalam JST diproses dalam neuron. Neuron-neuron tersebut terkumpul di dalam lapisan-lapisan yang disebut neuron layers. Lapisan-lapisan penyusun JST tersebut dapat dibagi menjadi :

- a. Lapisan *Input*, unit-unit di dalam lapisan input disebut unit-unit input. Unit-unit input tersebut menerima pola inputan data dari luar yang menggambarkan suatu permasalahan.
- b. Lapisan Tersembunyi, unit-unit di dalam lapisan tersembunyi disebut unit-unit tersembunyi. Dimana outputnya tidak dapat secara langsung diamati.
- c. Lapisan *Output*, unit-unit di dalam lapisan output disebut unit-unit *output*. *Output* dari lapisan ini merupakan solusi JST terhadap suatu permasalahan.

2.6 Self Organizing Maps (SOM)

Menurut (Munawar, 2015) Kohonen *Self Organizing Map* (SOM) merupakan salah satu algoritma *clustering* yang paling populer dan merupakan salah satu *tool* visualisasi yang handal untuk memproyeksikan hubungan kompleks dari ruang *input* berdimensi tinggi kedalam sebuah ruang berdimensi rendah (biasanya berupa grid 2 dimensi). Selama proses penyusunan diri, *cluster* yang memiliki *vector* bobot paling cocok dengan pola *input* (memiliki jarak yang paling dekat) akan terpilih sebagai pemenang. Selanjutnya, neuron yang menjadi pemenang beserta dengan neuron-neuron tetangganya akan memperbaiki bobot-bobotnya masing-masing (Nafi'iyah, 2015). Arsitektur jaringan Kohonen SOM seperti Gambar 2.5.



Gambar 2.5 Arsitektur Jaringan Kohonen SOM (Munawar, 2015)

Langkah atau tahapan algoritma *Self Organizing Maps* (SOM) adalah :

- a. Inisialisasi vektor *input* $x_1, x_2, x_3, \dots, x_n$

- b. Inisialisasi neuron *output* sebanyak $y_1, y_2, y_3, \dots, y_n$
- c. Menentukan *weight* (bobot) neuron output dengan nilai antara x_{min} dan x_{max}
- d. Mengulangi langkah e sampai h hingga tidak ada update *weight* (bobot) atau telah mencapai kondisi berhenti (*error* terkecil)
- e. Pemilihan acak salah satu data dari vektor *input* sebagai data *training*
- f. Mencari jarak terdekat dari masing-masing neuron output ke data *input* menggunakan rumus euclidian distance. Perhitungan jarak menggunakan Persamaan (3).

$$De = \sqrt{(xi - si)^2 + yi - ti^2} \dots\dots\dots(3)$$

Dari seluruh bobot (D_i) dicari yang paling kecil jaraknya, indeks dari bobot (D_i) ini disebut *winning* neuron

- g. Untuk setiap bobot w_{ij} diperbaharui bobot tetangga menggunakan rumus dengan persamaan (4).

$$w_{ij}(baru) = w_{ij}(lama) + \alpha[x_i - w_{ij}(lama)] \dots\dots\dots(4)$$

Keterangan :

$w_{ij}(baru)$ = bobot matriks baru

$w_{ij}(lama)$ = bobot matriks lama

α = *learning rate*/alfa

x_i = nilai setiap vektor *input*

- h. Mengupdate bobot bias (*error*) *learning rate*
- i. Simpan bobot yang telah konvergen

2.7 Perbandingan SOM dan K-Means

Perbandingan algoritma *Self Organizing Maps* (SOM) dan K-Means adalah sebagai berikut :

a. Nature of Clustering

SOM adalah algoritma *unsupervised learning* yang memiliki unsur pelatihan, sedangkan *K-Means* adalah algoritma *clustering* biasa.

b. Representasi Data

SOM membuat representasi visual dari data sebagai peta, sedangkan *K-Means* hanya memberikan hasil *clustering* berupa angka.

c. Jumlah Clusters

SOM memiliki jumlah *cluster* yang lebih fleksibel dibandingkan *K-Means* yang memiliki jumlah *cluster* tetap.

d. Dimensionalitas

SOM lebih baik digunakan untuk data dengan dimensionalitas tinggi, sedangkan *K-Means* lebih baik digunakan untuk data dengan dimensionalitas rendah.

e. Keandalan Hasil

K-Means memiliki hasil yang lebih stabil dan konsisten dibandingkan SOM, tetapi SOM memberikan representasi visual yang lebih baik dari data.

f. Proses Pelatihan

SOM memerlukan waktu pelatihan yang lebih lama dibandingkan *K-Means*, tetapi hasil dari SOM lebih baik dalam memvisualisasikan pola dalam data.

g. Konsep Dasar

SOM berdasarkan pada konsep pembelajaran dan representasi data, sedangkan *K-Means* berdasarkan pada konsep optimisasi dalam pengelompokan data.

2.8 Simpangan Baku

Menurut cara validasi untuk menghasilkan kualitas pengelompokan terbaik dilakukan dengan memperhatikan nilai rasio rata-rata simpangan baku dalam kelompok dan simpangan baku antar kelompok. Nilai variansi kelompok adalah hasil dari pembagian S_w dengan S_b dimana nilainya semakin kecil menunjukkan bahwa semua *cluster* semakin baik. Rata-rata simpangan baku dan rata-rata di dalam kelompok S_w simpangan baku antar kelompok S_b dinyatakan dengan perhitungan menggunakan Persamaan (5) dan (6).

$$s_w = \frac{1}{c} \sum_{k=1}^c s_k \quad \dots\dots\dots(5)$$

Keterangan :

- s_w = rata-rata simpangan baku dalam kelompok
- s_k = simpangan baku tiap *cluster*
- c = jumlah *cluster*

$$s_b = \left[\frac{1}{c-1} \sum_{k=1}^c (\bar{X}_k - \bar{X})^2 \right]^{\frac{1}{2}} \quad \dots\dots\dots(6)$$

Keterangan :

- s_b = rata-rata simpangan baku antar kelompok

- c = jumlah *cluster*
 \bar{X}_k = rata-rata tiap cluster
 \bar{X} = rata-rata keseluruhan

Metode pengelompokan yang baik memiliki nilai simpangan baku antar kelompok (S_b) yang maksimum dan nilai simpangan baku dalam kelompok (S_w) yang minimum. Nilai rasio S_w terhadap S_b yang semakin kecil dapat dikatakan metode tersebut mempunyai kinerja yang baik karena memiliki tingkat homogenitas yang tinggi di dalam cluster (Satria & Aziz , 2016). Rasio S_w terhadap S_b dapat dihitung dengan rumus Persamaan (7).

$$S = \frac{S_w}{S_b} \dots \dots \dots (7)$$

2.9 Penelitian Terkait

Adapun penelitian terdahulu yang terkait dengan penelitian saat ini dipaparkan pada tabel 2.1.

Tabel 2.1 Penelitian Terkait

No	Judul, Penulis, Tahun	Algoritma	Dataset	Akurasi
1	Ketepatan Klasifikasi Penerima Beasiswa STIMIK STIKOM Bali dengan Hybird <i>Self Organizing Maps</i> dan Algoritma <i>K-Means</i> (Hendayanti, Putri and Nurhidayati, 2018)	<i>Self Organizing Maps</i> dan <i>K-Means</i>	Data penerima beasiswa tahun 2017	SOM=54,45% K-Means=54,45%
2	Penggunaan <i>Self Organizing Map</i> Dalam Pengelompokan Tingkat Kesejahteraan Masyarakat (Irwan <i>et al.</i> , 2020)	<i>Self Organizing Maps</i>	Data wilayah Kabupaten/Kota di Provinsi Sulawesi Selatan berdasarkan	Model pembentukan dengan 5 cluster merupakan cluster terbaik dengan nilai DBI

No	Judul, Penulis, Tahun	Algoritma	Dataset	Akurasi
			statistik kunci Kesejahteraan Rakyat Provinsi Sulawesi Selatan Tahun 2018	terkecil yaitu 0,173.
3	Identifikasi Calon Penerima Bantuan Satu Keluarga Satu Sarjana (SKSS) Menggunakan Algoritma K-Means (Febrianti and Lubis, 2022)	<i>K-Means</i>	Data calon penerima bantuan SKSS Kecamatan Tiumang	Terdapat 8 keluarga yang layak menerima bantuan dengan nilai 32%
4	Perbandingan Kinerja Metode Ward dan K-Means dalam Menentukan Cluster Data Mahasiswa Pemohon Beasiswa (Studik Kasus: STIMIK Pringsewu) (Satria and Aziz, 2016)	<i>Ward</i> dan <i>K-Means</i>	Data calon penerima beasiswa PPA dan BBA	Nilai PPA Ward dan K-Means = 0,749959584 % Nilai BBA ward = 0.5346668 % K-Means = 0,831525302 %
5	Perbandingan Algoritma <i>K-Means</i> dan <i>SFCM</i> pada Pengelompokan Rumah Tangga Miskin (Widayani and Harliana, 2020)	<i>K-Means</i> dan <i>Fuzzy Subtractive Clustering</i> (SFCM)	Data hasil DDRT Dinas Ketahanan Pangan Kabupaten Cirebon	Nilai akurasi yang dihasilkan oleh SFCM lebih besar bila dibandingkan dengan K-Means yaitu 94,8%