

BAB II TINJAUAN PUSTAKA

A. Ekstrakurikuler

Ekstrakurikuler merupakan kegiatan pelajaran tambahan dan kegiatan murid yang dilakukan di sekolah, tidak sebagai sekedar tambahan atau sebagai kegiatan yang berdiri sendiri. Sedangkan orientasi kegiatan ekstrakurikuler adalah untuk memperkaya dan memperluas wawasan keilmuan dan kepribadian serta meningkatkan kemampuan tentang sesuatu yang telah dipelajari dalam satu bidang studi (Sutisna, 1983).

Melalui kegiatan ekstrakurikuler siswa dapat bertambah wawasan mengenai mata pelajaran yang erat kaitannya dengan pelajaran di ruang kelas dan biasanya yang membimbing siswa dalam mengikuti ekstrakurikuler adalah guru bidang studi yang bersangkutan. Melalui kegiatan ekstrakurikuler juga siswa dapat menyalurkan bakat, minat, dan potensi yang dimiliki. Salah satu ciri kegiatan ekstrakurikuler adalah keanekaragamannya, hampir semua minat remaja dapat digunakan sebagai bagian dari kegiatan ekstrakurikuler.

B. Machine Learning

Machine learning merupakan salah satu dari disiplin ilmu dari Kecerdasan Buatan (*Artificial Intellegence*). Pendekatan *machine learning* dapat digunakan untuk menyelesaikan masalah-masalah kompleks di berbagai bidang yang tidak dapat diselesaikan dengan program konvensional, misalnya seperti deteksi penyakit di bidang kesehatan, program komputer untuk mengajar di bidang pendidikan, pengenalan tulisan, pencarian informasi, analisis pasar, dsb. *Machine learning* dapat dikategorikan ke dalam kelompok *unsupervised learning* dan *supervised learning* (Santoso dan Azis, 2020).

1. *Supervised Learning*

Supervised Learning adalah metode pembelajaran dengan pengawasan. Metode ini berusaha menemukan hubungan antara input dengan output, dengan demikian jenis metode pembelajaran ini memiliki variabel output (bebas) dan variabel input (terikat), sehingga ada hubungan sebab akibat (Santoso dan Azis, 2020).

2. *Unsupervised Learning*

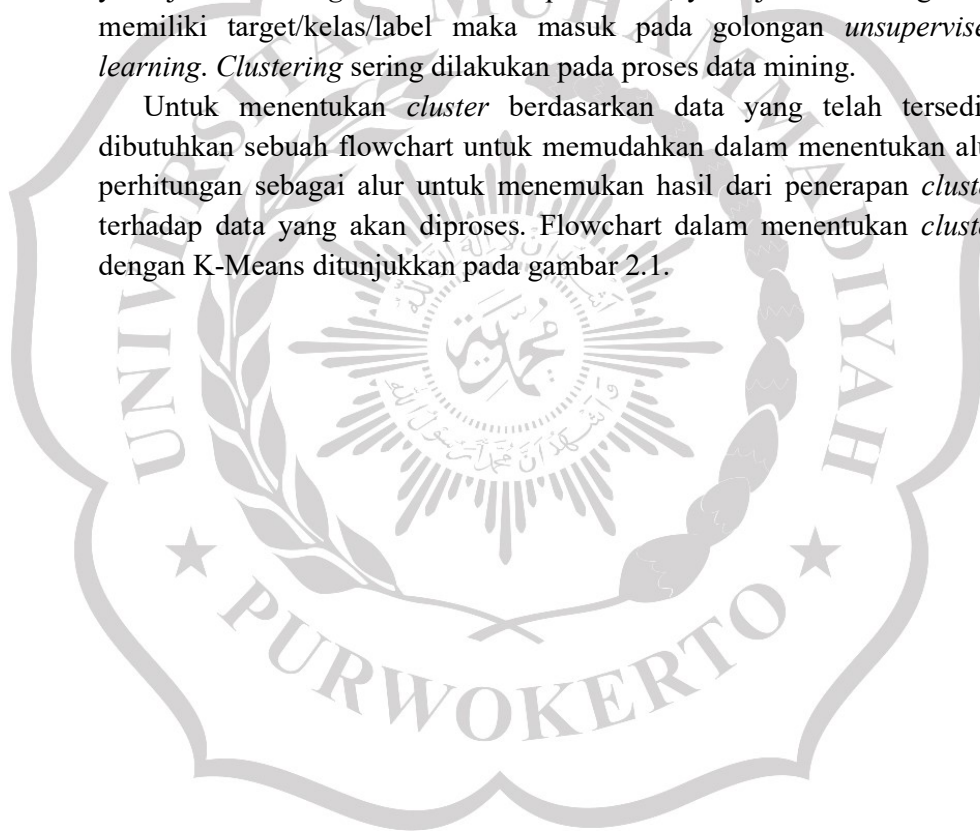
Unsupervised Learning adalah metode pembelajaran tanpa pengawasan. Metode ini berusaha menemukan hubungan antar variabel

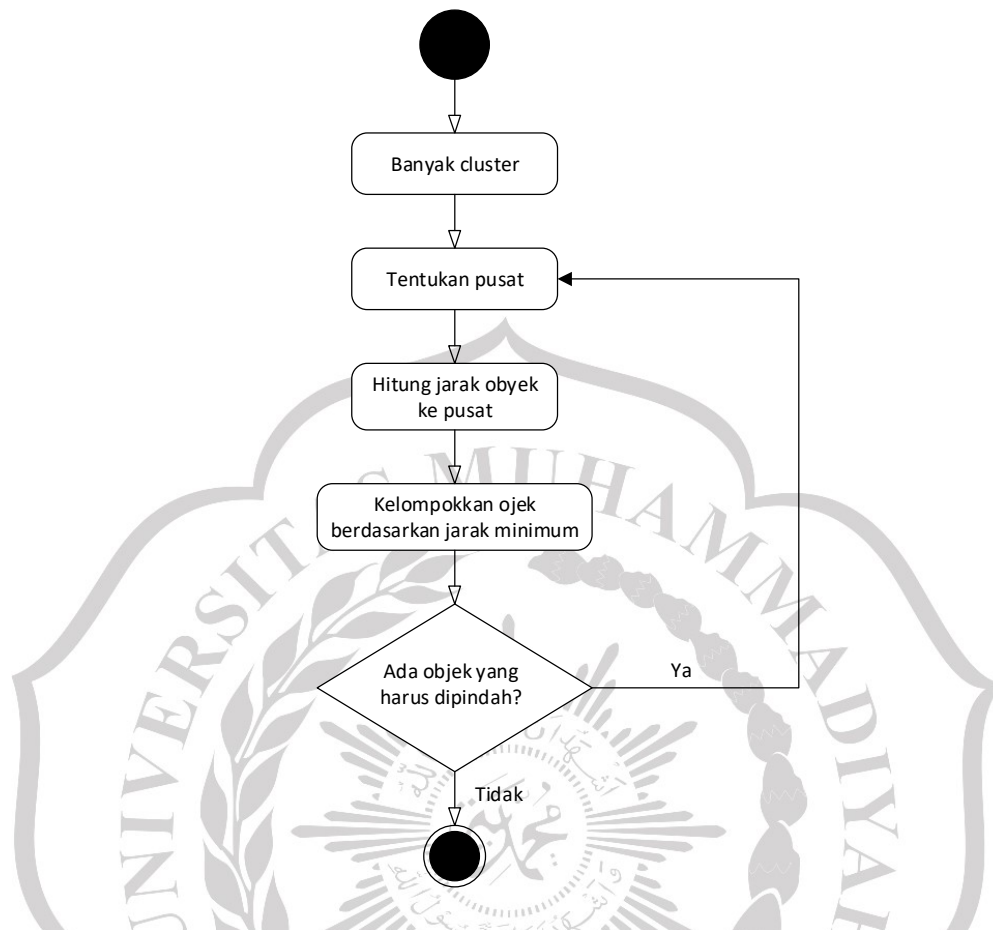
bebas, atau hubungan antar data pada suatu variabel, atau menghitung nilai statistik tertentu dari satu atau lebih variabel bebas. Dengan demikian, jenis metode pembelajaran ini tidak memiliki variabel output (Santoso dan Azis, 2020).

C. *Clustering*

Menurut Chintia Devi (2020), *Clustering* adalah proses dari pengelompokan data yang telah dianalisa pada suatu kasus kedalam kelas-kelas yang dianggap mirip. Pada setiap *cluster* memiliki data yang karakteristiknya sama, tetapi memiliki karakteristik yang berbeda dengan data dari *cluster* lainnya. *Clustering* dan klasifikasi memiliki perbedaan, yaitu jika *clustering* tidak memiliki perbedaan, yaitu jika *clustering* tidak memiliki target/kelas/label maka masuk pada golongan *unsupervised learning*. *Clustering* sering dilakukan pada proses data mining.

Untuk menentukan *cluster* berdasarkan data yang telah tersedia, dibutuhkan sebuah flowchart untuk memudahkan dalam menentukan alur perhitungan sebagai alur untuk menemukan hasil dari penerapan *cluster* terhadap data yang akan diproses. Flowchart dalam menentukan *cluster* dengan K-Means ditunjukkan pada gambar 2.1.





Gambar 2.1 Flowchart K-means (Hermawati, 2013)

D. *K-Means Clustering*

Menurut Nainggolan dan Lumbantoruan (2018), *K-Means* adalah salah satu metode dari *Machine Learning* yang memiliki logika paling sederhana dan paling populer. *K-Means clustering* ini bertujuan untuk melakukan pengelompokan bersama dalam satu set objek dengan menggunakan cara dimana objek yang ada atau sama atau mirip dalam *cluster* dibandingkan dengan objek *cluster* lain. Metode ini berfungsi untuk mempartisi data ke dalam *cluster* atau kelompok sehingga data yang memiliki keunikan atau karakteristik yang sama akan dikelompokkan dalam satu *cluster* yang sama dan jika data lain yang mempunyai karakteristik atau kesamaan yang berbeda akan dikelompokkan ke dalam kelompok data yang baru (Pusvitaningrum, 2020).

Algoritma *K-Means clustering* merupakan salah satu algoritma pengelompokan data dengan sistem partisi. Untuk itu digunakan aturan dalam Algoritma *K-Means* sebagai berikut :

- Jumlah *cluster* atau *k* harus diinisialisasikan terlebih dahulu
- Atribut bersifat numerik
- Keterbatasan atribut
- Kompleksitas algoritma linear (*n*) (Nainggolan dan Lumbantoruan, 2018).

Algoritma *K-Means* termasuk dalam metode *non-hierarchical* yang mempartisi data dalam bentuk satu atau lebih *cluster*, sehingga data yang mempunyai karakteristik yang sama dikelompokkan dalam satu *cluster* yang sama dan data yang memiliki karakteristik berbeda dikelompokkan ke dalam *cluster* lain. Algoritma ini merupakan algoritma yang paling umum digunakan karena mudah untuk diimplementasikan (Nainggolan dan Lumbantoruan, 2018).

Berikut ini urutan Algoritma *K-Means*:

- Menentukan jumlah *k-cluster* yang akan dibentuk
- Membangkitkan *k-centroid* (titik pusat *cluster*) secara acak (Nainggolan dan Lumbantoruan, 2018).

Untuk menghitung jarak setiap data terhadap masing-masing *centroid* digunakan rumus jarak Euclidean (Euclidean Distance) dengan persamaan (1) sebagai berikut:

$$\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \dots \dots \dots (1)$$

Keterangan :

x_1 = objek *x* ke-1

x_2 = objek *x* ke-2

y_1 = data *y* ke-1

y_2 = data *y* ke-2

Ukuran jarak atau ketidaksamaan antar objek ke-i dengan objek ke-j, disimbolkan dengan d_{ij} dan $k=1, \dots, p$. nilai d_{ij} diperoleh melalui perhitungan jarak kuadrat Euclidean yang ditunjukkan pada persamaan (2) sebagai berikut:

$$d_{ij} = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} \dots \dots \dots (2)$$

Keterangan :

- d_{ij} = jarak kuadrat Euclidean antar objek ke-i dengan objek ke-j
- p = jumlah variabel ke-j
- x_{ik} = nilai atau data dari objek ke-i pada variabel ke-k
- x_{jk} = nilai atau data dari objek ke-j pada variabel ke-k

E. Metode *Elbow*

Metode *Elbow* adalah suatu metode sederhana yang sering digunakan untuk menghasilkan informasi sederhana, dimana biasanya sering digunakan untuk menentukan jumlah *cluster* atau titik *centroid* terbaik dengan cara melihat persentase dari hasil perbandingan antara jumlah *cluster* dari data yang ada, dimana dari data tersebut akan membentuk siku pada suatu titik atau biasa disebut dengan titik siku (Pusvitaningrum, 2020).

Menurut Sulistiyawan dan Hapsery (2021) Untuk mendapatkan perbandingan antara jumlah *cluster* yang akan membentuk siku pada suatu titik yaitu dengan menghitung SSE (*Sum of Square Error*) dari masing-masing nilai *cluster*. Semakin besar jumlah *cluster* K maka nilai SSE akan semakin kecil. Rumus SSE pada K-Means ditunjukkan pada persamaan (3) yaitu :

$$SSE = \sum_{k=1}^K \sum_{x_i \in S_k} \|X_i - C_k\|_2^2 \dots \dots \dots (3)$$

Keterangan :

- k = *cluster*
- X_i = data ke-i dalam *cluster* x
- C_k = nilai rata-rata k

Setelah dilihat akan ada beberapa nilai K yang mengalami penurunan paling besar dan selanjutnya hasil dari nilai K akan turun secara perlahan-lahan sampai hasil dari nilai K tersebut stabil (Muningsih dan Kiswati, 2018).

Algoritma Metode *Elbow* dalam menentukan nilai K pada K-Means

1. Mulai
2. Inisialisasi awal nilai K
3. Naikkan nilai K
4. Hitung hasil *sum of square error* dari nilai K
5. Melihat hasil *sum of square error* dari nilai K yang turun secara drastis
6. Tetapkan nilai K yang berbentuk siku
7. Selesai (Muningsih dan Kiswati, 2018).

