

## **BAB II**

### **TINJAUAN PUSTAKA**

#### **2.1. Penelitian Terkait**

Dalam penulisan penelitian ini penulis akan mencoba mengaitkan dengan beberapa penelitian terdahulu, supaya akan mendapatkan keterkaitan dengan penelitian diatas. Adapun penelitian terdahulu yang dimaksud oleh penulis adalah sebagai berikut:

Penelitian Fauziah Nur, Prof. M. Zarlis, dan Dr. Benny Benyamin Nasution yang berjudul: “Penerapan Algoritma *K-Means* Pada Siswa Baru Sekolah Menengah Kejuruan Untuk *Clustering* Jurusan” tahun 2017. Penelitian ini membahas tentang penerapan algoritma *K-Means* untuk pengelompokan data siswa sesuai kriteria-kriteria yang telah diambil dari data siswa. Adapun hasil yang diperoleh dari penelitian ini adalah pengelompokan data siswa yang menunjukkan jurusan apa yang paling banyak diambil yaitu rekayasa perangkat lunak dan berapa jumlah siswa yang tidak lulus(Nur et al., 2017).

Penelitian Febrizal Alfarasy Syam yang berjudul: “Implementasi Metode Klastering *K-Means* Untuk Mengelompokan Hasil Evaluasi Mahasiswa” tahun 2017. Penelitian ini membahas tentang penerapan algoritma *K-Means* untuk pengelompokan mahasiswa dengan mengklaster data akademik mahasiswa menjadi empat buah klaster, yaitu klaster mahasiswa berprestasi, berpotensi berprestasi, berpotensi bermasalah, dan klaster mahasiswa bermasalah. Adapun hasil yang diperoleh dari penelitian ini adalah pengelompokan data akademik mahasiswa dapat berfungsi sebagai acuan bagi perencana akademik untuk memantau dan mengevaluasi perkembangan kinerja akademik setiap mahasiswa(Syam, 2017).

Penelitian Agil Aditya, Ivan Jovian, Betha Nurina Sari yang berjudul: “Implementasi *K-Means Clustering* Ujian Nasional Sekolah Menengah Pertama di Indonesia Tahun 2018/2019” tahun 2020. Penelitian ini membahas tentang penerapan algoritma *K-Means* untuk *clustering* menggunakan data capaian Ujian Nasional Sekolah Menengah Pertama pada tahun 2018/2019 untuk mengelompokan provinsi mana saja di Republik Indonesia yang memiliki nilai

Ujian Nasional tingkat Sekolah Menengah Pertama yang rendah, sedang dan tinggi. Adapun hasil yang diperoleh dari penelitian ini adalah data yang menunjukkan pengelompokan hasil Ujian Nasional sehingga nantinya dapat dijadikan bahan evaluasi terhadap hasil ujian di tiap-tiap provinsi (Aditya et al., 2020).

Jadi dari tinjauan pustaka tersebut dapat ditemukan titik-titik persamaan dan juga perbedaan dengan penelitian yang penulis teliti. Adapun titik persamaannya yaitu sama-sama menggunakan algoritma *K-Means* dan titik perbedaannya yaitu terletak di penerapan algoritma *K-Means* yang dimana tempat penggunaannya yang berbeda-beda.

## **2.2. Bimbingan Konseling**

Bimbingan adalah suatu proses pemberian bantuan yang terus menerus dan secara sistematis pada individu dalam memecahkan masalah yang sedang dihadapi individu tersebut, supaya tercapai kemampuan untuk memahami dirinya sendiri (*self understanding*), kemampuan untuk menerima dirinya sendiri (*self acceptance*), kemampuan untuk mengarahkan dirinya sendiri (*self direction*), dan kemampuan untuk merealisasikan dirinya sendiri (*self realization*) sesuai dengan potensi dan kemampuannya dalam mencapai penyesuaian diri dengan lingkungannya sendiri, baik keluarga, sekolah dan masyarakat (Surya & Djumhur, 1975). Rochman Natawidjaja berpendapat bahwa bimbingan adalah suatu proses pemberian bantuan kepada individu yang dilakukan secara berkesinambungan, agar individu tersebut dapat memahami dirinya sendiri, sehingga dapat mengarahkan dirinya dan dapat bertindak secara wajar, sesuai dengan tuntutan dan keadaan lingkungan sekolah, keluarga, masyarakat, dan kehidupan pada umumnya (Natawijaya, 1988). Menurut Wisnu Pramuja Utama bimbingan adalah proses bantuan yang terus menerus kepada individu untuk mencapai kemampuan untuk dapat memahami dirinya sendiri dan kemampuan untuk merealisasikan dirinya sendiri sesuai dengan potensi atau kemampuannya dalam mencapai penyesuaian diri dalam lingkungan, baik keluarga, sekolah, dan masyarakat, yang pada akhirnya individu itu akan mencapai perkembangan yang optimal.

Konseling adalah suatu pertemuan langsung dengan individu yang diutamakan pada pemberian bantuan kepada individu tersebut sehingga dapat

menyesuaikan dirinya sendiri secara lebih efektif dengan (McDaniel, 1956). Menurut Tolbert konseling adalah hubungan pribadi yang dilakukan secara tatap muka antara dua individu yang mana konselor melalui hubungan itu dengan kemampuan-kemampuan khusus yang dimilikinya menyediakan situasi belajar. Dalam hal ini konseli dibantu untuk dapat memahami diri sendiri, keadaannya sekarang, dan kemungkinan keadaannya dimana depan yang dapat ia ciptakan dengan menggunakan potensi yang dimilikinya, demi untuk kesejahteraan pribadi maupun masyarakat. Lebih lanjut konseli dapat belajar bagaimana memecahkan masalah-masalah dan menemukan kebutuhan-kebutuhan yang akan datang (Prayitno, 2004). Jones menjelaskan bahwa konseling merupakan suatu hubungan profesional antara seorang konselor yang terlatih dengan klien. Hubungan ini biasanya bersifat individual atau seorang-seorang, meskipun terkadang melibatkan lebih dari dua orang dan dirancang untuk membantu klien memahami dan memperjelas pandangan terhadap ruang lingkup hidupnya, sehingga dapat membuat pilihan yang bermakna bagi dirinya (Insano, 2004).

Dari penjelasan diatas maka penulis dapat menyimpulkan bahwa pengertian dari bimbingan konseling adalah suatu proses dari hubungan yang dilakukan secara tatap muka yang dilakukan oleh dua individu atau lebih yaitu seorang konselor dengan klien yang bersifat tertutup yang dimaksudkan untuk mengatasi permasalahan yang sudah, sedang, atau akan dialami oleh klien atau konseli.

Sifat bimbingan dan konseling menurut Andi Mappiere(Mappiare AT, 2002)dapat dijelaskan seperti dibawah ini:

1. Pencegahan(*preventive*).

Bimbingan dan konseling yang bersifat pencegahan artinya memberikan bantuan terutama pada murid sebelum murid menghadapi persoalan atau kesulitan yang serius.

2. Pengembangan(*development*).

Bimbingan dan konseling yang bersifat pengembangan artinya usaha bantuan yang diberikan kepada murid dengan seiring perkembangna mental murid tersebut terutama untuk menetapkan jalan berfikir dan bertindak murid sehingga nantinya akan dapat berkembang dengan optimal.

### 3. Penyembuhan(*curative*)

Bimbingan dan konseling yang bersifat penyembuhan artinya usaha bantuan yang diberikan kepada murid selama atau setelah murid mengalami masalah atau persoalan serius yang bermaksud agar murid tersebut terbebas dari masalah atau persoalan tadi.

### 4. Pemeliharaan(*treatment*)

Bimbingan dan konseling yang bersifat pemeliharaan artinya usaha bantuan yang dimaksudkan terutama untuk memupuk dan mempertahankan kesehatan mental dari murid yang bersangkutan dapat bertahan dalam kesembuhan setelah sudah menjalani proses penyembuhan.

## 2.3. *Data Mining*

Sebagai sebuah bidang ilmu yang bisa dianggap relatif baru pada saat ini, *data mining* menjadi salah satu pusat perhatian dari para praktisi maupun para akademisi. Menurut Suntoro *data mining* adalah suatu proses untuk mendapatkan informasi yang berguna dari sebuah dataset yang besar dan perlu untuk diekstraksi supaya menjadi informasi yang baru dan dapat digunakan untuk membantu dalam pengambilan keputusan (Suntoro, 2019). Menurut Witten *data mining* adalah sebuah proses untuk menganalisa data dan menyimpulkannya menjadi sebuah informasi atau pola yang penting untuk meningkatkan keuntungan, memperkecil biaya ataupun keduanya (Ian H. Witten, 2016).

*Data mining* adalah proses yang memakai teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk melakukan ekstraksi dan identifikasi informasi yang bermanfaat dan juga pengetahuan terkait dengan berbagai dataset besar (Turban, 2005). Menurut Han(Han, 2012) *data mining* adalah proses untuk menemukan pola yang menarik dari data yang berjumlah besar.

## 2.4. *Preprocessing*

*Preprocessing* adalah suatu proses yang dilakukan terhadap data sebelum data diproses dengan algoritma data mining, *preprocessing* ini meliputi integrasi, seleksi, data cleaning, dan transformasi (Gunadi & Sensuse, 2012).

Dari data siswa dilakukan *preprocessing* yang bersifat transformasi dengan mencari fscore dan fabs lalu melakukan normalisasi fitur dari fscore dan fabs yang sudah didapatkan. Fscore adalah hasil *preprocessing* yang didapatkan dari rata-rata data nilai pengetahuan dan nilai keterampilan, dan fabs adalah hasil *preprocessing* yang didapatkan dari pengurangan total jumlah hari efektif yang dikurangi dengan total absensi siswa.

$$Fscore = \frac{Peng+Ket}{2} \quad (1)$$

Dimana

Peng: nilai pengetahuan.

Ket: nilai keterampilan.

$$Fabs = n - (S + I + A) \quad (2)$$

Dimana

n: jumlah hari efektif.

S: total siswa tidak masuk karena sakit.

I: total siswa tidak masuk karena izin.

A: total siswa tidak masuk tanpa keterangan.

Normalisasi adalah suatu teknik untuk pemetaan atau penskalaan atau *preprocessing*. Normalisasi memiliki berbagai macam teknik seperti *Min-Max*, *Z-Score*, dan *feature scaling*. Penelitian ini menggunakan teknik normalisasi *Min-Max*. Normalisasi *Min-Max* adalah metode normalisasi yang sering dipakai untuk mengatasi permasalahan antar fitur yang mempunyai jarak yang terlalu jauh (Ramadhan & Khoirunnisa, 2021). Normalisasi dilakukan pada fitur fscore dan fabs sehingga didapatkan nilai fscore dan fabs diskalakan kedalam range 0 sampai 1 dengan menggunakan rumus sebagai berikut:

$$N^* = \frac{N - \min(n)}{\max(n) - \min(n)} \quad (3)$$

Dimana:

$N^*$  : hasil normalisasi.

$N$  : data yang belum dinormalisasi.

$\max(n)$ : nilai maksimum dari semua data.

$\min(n)$  : nilai minimum dari semua data.

## 2.5. *Clustering*

*Clustering* adalah suatu metode pengelompokan data kedalam beberapa *cluster* atau kelompok sehingga data yang ada dalam satu *cluster* mempunyai tingkat kemiripan yang maksimal sedangkan data antar *cluster* yang berbeda mempunyai tingkat kemiripan yang minimal (Tan, 2006). Menurut Madhu Yedha *clustering* adalah proses pengorganisasian objek yang berupa data-data kedalam set kelas yang saling berhubungan satu dengan yang lainnya yang disebut *cluster*. *Clustering* adalah contoh dari klasifikasi yang tidak mempunyai arah atau disebut *unsupervised*. *Unsupervised* berarti pengelompokan tidak bergantung pada standard kelas ataupun pelatihan atau training (Gupitha, 2018).

*Clustering* pada data yang bersifat kategorial sudah mulai berkembang walaupun masih kalah jauh jika dibandingkan dengan *clustering* pada data yang bersifat numerik. Data kategorial secara alami tidak bisa diperlakukan sebagaimana data numerik karena beberapa operasi dalam data numerik yang tidak bisa untuk dilakukan dalam data kategorial seperti *mean* dan *median*. *Clustering* adalah alat utama dalam banyak aplikasi dalam analisa data statistik, *data mining*, information retrieval, pengolahan citra, dan lain sebagainya. *Clustering* bertujuan untuk mengelompokkan objek-objek yang berupa data kedalam beberapa *cluster* atau kelompok sehingga objek-objek yang berbentuk data tadi yang berada dalam satu *cluster* yang sama akan memiliki tingkat kemiripan yang tinggi sedangkan data yang berada di klaster yang berbeda akan memiliki tingkat kemiripan yang rendah. Inti dari *clustering* adalah untuk mempartisi suatu set objek dalam menguraikan

*cluster* homogen, sehingga nantinya objek milik *cluster* yang sama akan mirip satu dengan lainnya dibandingkan objek yang masuk kelompok yang berbeda. Berdasarkan pendekatan yang berada dalam penerapan keanggotaan dalam *cluster*, metode *clustering* secara umum dibagi jadi dua yaitu *hard clustering* yang tiap objeknya hanya dialokasikan kedalam suatu *cluster* baik selama operasi *clustering* ataupun dalam output *clustering* tersebut, dan *fuzzy clustering* yang selama operasi *clustering* berlangsung setiap objek dialokasikan kedalam beberapa *cluster* dan diberi derajat kanggotaan dengan nilai antara 0 dan 1. Output *fuzzy clustering* dapat diubah menjadi *hard clustering* dengan cara memilih nilai keanggotaan yang paling tinggi(Tan, 2006).

## **2.6. *K-Means Clustering***

Algoritma *K-Means* adalah salah satu algoritma yang bersifat *partitional* karena *K-Means* berdasarkan pada penentuan jumlah awal kelompok dengan mendefinisikan nilai centroidnya terlebih dahulu (Madhulatha, 2012). Algoritma *K-Means* mengulang-ulang proses secara terus menerus sampai mendapat basis data *cluster*. Diperlukan jumlah *cluster* awal yang diinginkan sebagai input dan menghasilkan titik *centroid* akhir sebagai output. *K-Means* akan memilih pola *K* untuk titik awal *centroid* secara random. Jumlah iterasi untuk mendapatkan *cluster centroid* akan dipengaruhi juga oleh calon *cluster centroid* awal yang sudah ditentukan secara random sehingga didapatkan suatu cara dalam pengembangan algoritma dengan cara menentukan *centroid cluster* yang dilihat dari kepadatan data awal yang tinggi supaya didapatkan kinerja yang lebih tinggi juga(Eltibi & Ashour, 2011).

Algoritma *K-Means* dalam penyelesaiannya nantinya akan menghasilkan titik *centroid* yang dapat dijadikan tujuan dari algoritma *K-Means*. Setelah iterasi berhenti setiap objek dalam dataset akan menjadi anggota dari suatu *cluster*. Nilai dari *cluster* ditentukan dengan mencari seluruh objek untuk menemukan *cluster* yang paling dekat jaraknya dengan objek. Algoritma *K-Means* melakukan pengelompokan data-data kedalam suatu dataset ke sebuah *cluster* yang paling dekat jaraknya (Bangoria, 2013). Nilai dari *centroid* awal yang sudah dipilih dengan random akan menjadi titik pusat awal, akan dihitung jaraknya dengan semua data

yang ada dengan menggunakan rumus Euclidian Distance, data yang mempunyai jarak paling pendek dari *centroid* akan membuat satu *cluster*. Proses ini terus menerus diulangi sampai tidak lagi terjadi perubahan di semua *cluster* (Bhatia & Khurana, 2013).

Langkah-langkah didalam algoritma *K-Means* yaitu:

1. Menentukan jumlah *cluster* yang akan dipakai untuk mengelompokkan data dengan memperhatikan beberapa faktor seperti pertimbangan konseptual dan teoritis.
2. Menentukan *K centroid* awal secara acak dari objek-objek yang ada dengan banyak yang sesuai dengan banyaknya *cluster*, sedangkan untuk menentukan nilai *centroid* yang merupakan tahap dari iterasi digunakan rumus seperti dibawah:

$$v_m = \frac{\sum_{k=1}^n x_i}{n} \quad (4)$$

Dimana

$v_m$ : *centroid* cluster *i*

$n$  : jumlah objek anggota cluster

$x_i$  : objek ke-*i*

3. Menghitung jarak tiap objek ke masing-masing *centroid* dari masing-masing *cluster* dengan menggunakan rumus Euclidian Distance seperti dibawah:

$$d_e = \sqrt{(x_i - a)^2 + (y_i - b_i)^2} \quad (5)$$

Dimana

$d_e$ : Euclidean Distance

$i$  : banyaknya objek

( $x,y$ ) : koordinat objek

( $a,b$ ) : koordinat *centroid*

4. Alokasikan masing-masing objek kedalam *centroid* yang paling dekat dengan mengukur jarak dari objek ke *centroid cluster* terdekat.
5. Ulangi dari langkah 2 sampai tidak ada lagi perubahan *centroid* dan tidak ada objek yang berpindah *cluster*.

## 2.7. *Silhouette Coefficient*

*Silhouette Coefficient* adalah gabungan dari dua metode yaitu metode kohesi yang berguna untuk mengukur seberapa dekatkah relasi antar objek dalam suatu cluster, dan metode separasi yang berguna untuk mengukur seberapa jauhkah suatu cluster dengan cluster yang lain, jadi *Silhouette Coefficient* digunakan untuk melihat kualitas dan kekuatan cluster seberapa baik suatu objek ditempatkan didalam sebuah cluster. Berikut adalah tahapan perhitungan *Silhouette Coefficient* (Anggara et al., 2016):

Hitung rata-rata jarak dari sebuah objek dalam 1 cluster.

$$a(i) = \frac{1}{|A|-1} \sum_{j \in A, j \neq i} d(i, j) \quad (6)$$

Dimana

$i, j$  : objek dalam cluster A

$d(i, j)$  : jarak antara objek i dan j

Hitung rata-rata jarak dari objek i tersebut dengan semua objek di cluster lain dan diambil nilai terkecilnya.

$$d(i, C) = \frac{1}{|A|} \sum_{j \in C} d(i, j) \quad (7)$$

Dimana

$d(i, C)$  : jarak rata-rata objek dengan objek diluar cluster

Setelah mendapatkan nilai  $d(i, C)$  selanjutnya memilih jarak yang terkecil sebagai  $b(i)$ .

$$b(i) = \min_{C \neq A} d(i, j) \quad (8)$$

Nilai *Silhouette Coefficient* merupakan jumlah  $s(i)$  yang didapatkan dengan mengombinasikan  $a(i)$  dan  $b(i)$ .

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (9)$$

Dimana

$s(i)$  : *Silhouette Coefficient*

$a(i)$  : rata-rata jarak antara i dan objek lain di clusternya

$b(i)$  : rata-rata jarak antara i dan objek di cluster lainnya

$d(i, j)$  : jarak antara objek  $i$  dan  $j$

Dengan ukuran nilai *Silhouette Coefficient* sebagai berikut

**Tabel 2. 1. Nilai *Silhouette Coefficient***

No	<i>Silhouette Coefficient</i>	Keterangan
1	1	Perbedaan <i>Cluster</i> signifikan satu dengan lainnya dan berjarak cukup jauh.
2	0	Perbedaan <i>Cluster</i> tidak terlalu signifikan bahkan tidak memiliki perbedaan.
3	-1	Berarti <i>cluster</i> diclusterkan dengan cara yang salah.