

BAB II

TINJAUAN PUSTAKA

II.1. Penelitian Terkait

Berdasarkan penelitian yang dilakukan oleh luluk Elvitaria (2017) dengan judul “Memprediksi Tingkat Peminat Ekstrakurikuler Pada Siswa SMK Analisis Kesehatan Abdurrah Menggunakan Algoritma C45 (Studi Kasus : SMK Analisis Kesehatan Abdurrah)” Pada penelitian ini menggambarkan proses data mining sebagai alat yang sederhana dan efisien dalam memonitor perkembangan pada tingkat peminat bahasa asing siswa-siswi meliputi 5 bahasa yaitu arab, inggris, jerman, mandarin, jepang. Dengan bantuan proses data mining, peneliti menggunakan metode Algoritma C4.5 dengan melakukan perhitungan Entropy dan Gain dari variabel hasil kuesioner yang dijawab siswa. Perbedaan terdahuludengan penelitian penulis, dimana data penelitian terdahulu memperdiksi peminat bahasa asing dihasilkan survei ke siswa dengan memberikan kuesioner yang akan dijawab, Sedangkan penulis, data yang diambil dari data tes penilaian siswa pada masing-masing bidang ekstrakurikuler dan kemudian dikelompokan kedalam k-means sehingga kita dapat mengetahui hasil pengelompokan data bidang ekstrakurikuler siswa pada SMK Negeri 1 Panai Hilir.

Penelitian yang dilakukan Rino Gupitha (2018) dengan judul “Penentuan Ekstrakurikuler (Eskul) SMK Terpadu Lampung Subang Dengan Algoritma C4.5” Pada Penelitian ini menggunakan pendekatan klasifikasi data mining untuk dapat menentukan ekstra kurikuler yang sesuai dan akan dipilih oleh siswa. Tujuan dari

klasifikasi adalah untuk menemukan model dari training set yang membedakan atribut ke dalam kategori atau kelas yang sesuai, kemudian digunakan untuk mengklasifikasi atribut yang kelasnya belum diketahui sebelumnya. Perbedaan penelitian terdahulu dengan penelitian penulis, dimana data penelitian terdahulu data berupa atribut Nama, Nilai Tes Tulis (TT), Nilai Tes Lisan (TTL), Nilai SKHUN), Sedangkan penulis, data yang diambil dari data tes penilaian siswa pada masing-masing bidang ekstrakurikuler dan kemudian dikelompokkan kedalam k-means sehingga kita dapat mengetahui hasil pengelompokan data bidang ekstrakurikuler siswa pada SMK Negeri 1 Panai Hilir.

Penelitian yang dilakukan sufajar Bustsianto (2019) berjudul “Penerapan Data Mining Terhadap Minat Siswa Dalam Mata Pelajaran Matematika Dengan Metode K – Means” Pada Penelitian ini menggambarkan algoritma clustering k – means sebagai pendukung keputusan untuk mengelompokkan minat siswa pada mata pelajaran matematika dengan metode algoritma k-means. Proses pengujian yang dilakukan oleh peneliti ini akurasi diukur dengan validasi DBI (Davies Bouldin Index) dan menghasilkan 2 cluster yaitu (minat) matematika dan (tidak minat) matematika. Perbedaan penelitian terdahulu dengan penelitian penulis, dimana data penelitian terdahulu menggunakan jumlah data set yang digunakan sebanyak 334 data kelas 10, 11 dan 12 dari semua jurusan di SMK Binamitra, Sedangkan penulis, data yang diambil dari data tes penilaian siswa pada masing-masing bidang ekstrakurikuler dan kemudian dikelompokkan kedalam k-means sehingga kita dapat mengetahui hasil pengelompokan data bidang ekstrakurikuler siswa pada SMK Negeri 1 Panai Hilir.

Penelitian yang dilakukan Eferoni Ndruru, dkk (2018) dengan judul “Implementasi Data Mining Dalam Pengelompokan Jurusan Yang Diminati Siswa SMK Negeri 1 Lolowa’u Menggunakan Metode K-Means Clustering” Pada penelitian ini Peneliti membangun / merancang sebuah Aplikasi yang dapat membantu pihak SMK Negeri 1 Lolowa’u untuk mengelompokkan data siswa pada jurusan yang diminati. Proses penelitian dilakukan berdasarkan data nilai siswa yang akan menentukan yang banyak memilih jurusan yang sama. Perbedaan penelitian terdahulu dengan penelitian penulis, dimana data penelitian terdahulu menggunakan jumlah dataset siswa yang akan ditentukan untuk memilih jurusan, Sedangkan penulis, data yang diambil dari data tes penilaian siswa pada masing-masing bidang ekstrakurikuler dan kemudian dikelompokkan kedalam k-means sehingga kita dapat mengetahui hasil pengelompokan data bidang ekstrakurikuler siswa pada SMK Negeri 1 Panai Hilir.

Berdasarkan penelitian yang dilakukan oleh Miftahul Hasanah, (2019) dengan judul “Implementasi Algoritma K-Means Untuk Klaterisasi Peserta Olimpiade Sains Nasional Tingkat SMA” Tujuan penelitian ini untuk membantu pihak sekolah dalam pemilihan Olimpiade Sains Nasional secara tepat dan efektif. Proses yang digunakan dalam penelitian ini adalah clustering dengan algoritma *K-Means* terhadap data nilai rapor siswa jurusan IPA di SMA Negeri 5 Sijunjung. Hasil penelitian ini mendapatkan 3 *cluster* siswa terhadap pemilihan Olimpiade Sains Nasional yaitu siswa yang sangat berkompeten, berkompeten dan kurang berkompeten. Perbedaan penelitian terdahulu dengan penelitian penulis, dimana data penelitian terdahulu menggunakan data nilai rapor akademik siswa

jurusan IPA di SMA Negeri 5 Sijunjung, Sedangkan penulis, data yang diambil dari data tes penilaian siswa pada masing-masing bidang ekstrakurikuler dan kemudian dikelompokkan kedalam k-means sehingga kita dapat mengetahui hasil pengelompokan data bidang ekstrakurikuler siswa pada SMK Negeri 1 Panai Hilir

II.2. Landasan Teori

Berikut ini akan dipaparkan mengenai landasan teori yang mendukung penelitian dan metode yang akan digunakan oleh penulis yang diambil dari beberapa jurnal dan buku.

II.2.1. Data Mining

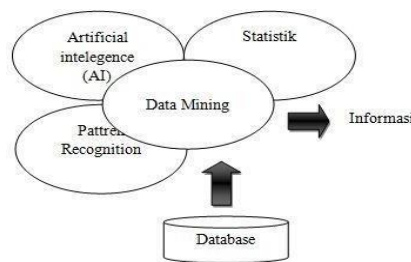
Data mining adalah proses yang mempekerjakan satu atau lebih teknik pembelajaran komputer (*machine computer*) untuk menganalisa dan mengekstrasi pengetahuan (*knowledge*) secara otomatis (Elmayati, 2017).

Data mining juga merupakan metode yang digunakan dalam pengolahan data berskala besar oleh karena itu data mining memiliki peranan yang sangat penting dalam beberapa bidang kehidupan diantaranya yaitu bidang industri, bidang keuangan, cuaca, ilmu dan teknologi. Dalam data mining juga terdapat metode – metode yang dapat digunakan seperti klasifikasi, clustering, regresi, seleksi variabel, dan market basket analisis (Elmayati, 2017).

Data mining juga bisa diartikan sebagai rangkaian kegiatan untuk menemukan pola yang menarik dari data dalam jumlah besar, kemudian data – data tersebut dapat disimpan dalam database, data warehouse atau penyimpanan informasi. Ada beberapa ilmu yang mendukung teknik data mining diantaranya

adalah data analisis, *signal processing*, *neural network* dan pengenalan pola (Elmayati, 2017).

Menurut Eko Prasetyo, Nama data mining sebenarnya mulai dikenal sejak tahun 1990. Munculnya data mining didasarkan pada jumlah data yang tersimpan dalam basis data semakin besar. Misalnya dalam sebuah supermarket, ada berapa transaksi pelanggan yang terjadi dalam sehari dan ada berapa juta data yang sudah tersimpan dalam sebulan. Dalam perusahaan, ada berapa juta data yang sudah tersimpan dari setiap kegiatan produksi untuk setiap produk yang dibuat dalam beberapa tahun (Teri Ade Putra, 2018).



Gambar II.1. Akar Ilmu Data Mining

Ada istilah lain yang mempunyai makna yang sama dengan data mining yaitu *knowledgediscovery in database* (KDD) bertujuan untuk memanfaatkan data dalam basis data dengan mengelolanya sehingga menghasilkan informasi baru yang berguna. Seperti dilustrasikan pada gambar II.1, jika dilacak akar keilmuannya, ternyata data mining mempunyai empat akar bidang ilmu sebagai berikut (Teri Ade Putra, 2018)

1. Statistik. Bidang ini merupakan akar paling tua, tanpa ada statistik maka data mining mungkin tidak ada. Dengan menggunakan statistik klasik ternyata data

yang diolah dapat diringkas dalam apa yang umum dikenal sebagai exploratory data analysis (EDA).

2. Kecerdasan buatan atau *artificial intelligence* (AI). Bidang ilmu ini berbeda dengan statistik. Teorinya dibangun berdasarkan teknik heuristik sehingga AI berkontribusi terhadap teknik pengolahan informasi berdasarkan pada model penalaran manusia.
3. Pengenalan pola. Sebenarnya data mining juga menjadi turunan bidang pengenalan pola, tetapi hanya mengolah data dari basis data. Data yang diambil dari basis data untuk diolah bukan dalam bentuk relasi, melainkan dalam bentuk normal pertama sehingga set data dibentuk menjadi bentuk normal pertama. Akan tetapi, data mining mempunyai ciri khas yaitu pencarian pola asosiasi dan pola sekuensial.
4. Sistem basis data. Akar bidang ilmu keempat dari data mining yang menyediakan informasi berupa data yang akan 'digali' menggunakan metode-metode yang disebut sebelumnya.

II.2.2. Clustering

Metode *cluster* yaitu menemukan kumpulan objek hingga objek - objek dalam satu kelompok sama (atau punya hubungan) dengan yang lain dan berbeda (atau tidak berhubungan) dengan objek - objek dalam kelompok lainnya. (Elmayati, 2017)

II.2.3. Algoritma K-Means

K-Means merupakan salah satu metode data *clustering non hirarki* yang berusaha mempartisi data yang ada ke dalam bentuk satu kelompok atau

lebih. Metode ini mempartisi data ke dalam *cluster* atau kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu cluster yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok yang lain. (Fitra Kurnia, 2019).

II.2.4. Langkah-langkah Metode

Langkah-langkah melakukan *Clustering* dengan metode *K-Means* adalah sebagaiberikut:

1. Pilih jumlah *Cluster* k sebagai jumlah *klaster* yang ingin dibentuk.
2. Inisialisasi ke pusat *Cluster* ini bisa dilakukan dengan berbagai cara. yang paling sering dilakukan adalah dengan random atau acak. Pusat-pusat *Cluster* diberi nilai awal dengan angka-angka random
3. Alokasikan semua data atau objek ke *Cluster* terdekat. Kedekatan dua objek ditentukan berdasarkan jarak ke-dua objek tersebut.

Untuk menghitung jarak setiap data ke pusat *cluster* menggunakan persamaan *Euclidean Distance*:

$$d(x,y) = \sqrt{\sum_{i=1}^n (X_i - Y_i)^2} \dots\dots\dots(1)$$

Keterangan:

$d(x,y)$ = jarak data ke x ke pusat *cluster* y

X_i = Koordinat dari objek X_i pada dimensi i

Y_i = Koordinat dari objek Y_i pada dimensi i

4. Kelompokkan setiap data berdasarkan jarak terdekat antara data dengan centroidnya.
5. Tentukan centroid baru dengan cara menghitung nilai rata-rata dari data-data

yang ada pada centroid yang sama dengan rumus seperti persamaan. Dimana n adalah banyaknya dokumen dalam cluster i dan x adalah data yang akan dihitung.

$$C_I = \left(\frac{\sum x}{n} \right) \dots \dots \dots (2)$$

6. Kembali ke langkah 3 jika posisi centroid baru dengan centroid lama tidak sama. (Teguh Wibowo, 2018)

II.2.5. Ekstrakurikuler

Kegiatan ekstrakurikuler merupakan sebuah kegiatan tambahan di sekolah yang pada umumnya dilaksanakan diluar jam pelajaran dan kegiatan ini bertujuan agar siswa lebih memperdalam dan mengembangkan apa yang dipelajari saat proses pembelajaran di kelas serta dapat mengembangkan minat dan bakat siswa. (Septiana Intan Pratiwi, 2020).

Ekstrakurikuler merupakan sebuah kegiatan tambahan yang diselenggarakan diluar jam pelajaran yang bertujuan untuk upaya pematapan kepribadian peserta didik. Kegiatan ekstrakurikuler adalah kegiatan yang dilakukan dalam mengembangkan aspek-aspek tertentu dari apa yang ditemukan pada kurikulum yang sedang dijalankan, termasuk yang berhubungan dengan bagaimana penerapan sesungguhnya dari ilmu pengetahuan yang dipelajari siswa sesuai dengan tuntutan kebutuhan hidup mereka maupun lingkungan sekitarnya. (Septiana Intan Pratiwi, 2020).

II.2.6. PHP (*Personal Hypertext Preprocessor*)

PHP (*Personal Hypertext Preprocessor*) adalah suatu bahasa pemrograman yang hanya dapat berjalan pada sisi server (*Server Side Scripting*), atau sebuah bahasa pemrograman yang digunakan untuk proses aksi yang terdapat dalam konten web. Artinya proses yang dibuat dengan php tidak akan berjalan tanpa menggunakan web server (Elmayanti, 2017).

Kelebihan PHP dari pemrograman lainnya adalah:

1. Bahasa pemrograman PHP adalah sebuah bahasa script yang tidak melakukan sebuah kompilasi dalam penggunaannya.
2. Web Server yang mendukung PHP dapat ditemukan dimana - mana dari mulai IIS sampai dengan apache, dengan konfigurasi yang relatif mudah
3. Dalam sisi pengembangan lebih mudah, karena banyaknya milis - milis dan developer yang siap membantu dalam pengembangan.
4. Dalam sisi pemahaman, PHP adalah bahasa scripting yang paling mudah karena referensi yang banyak.
5. PHP adalah bahasa open source yang dapat digunakan diberbagai mesin (linux, unix, windows) dan dapat dijalankan secara runtime melalui console serta juga dapat menjalankan perintah-perintah sistem.

II.2.7. UML (*Unified Modelling Language*)

Hasil pemodelan pada OOAD terdokumentasikan dalam bentuk *Unified Modeling Language* (UML). UML adalah satu alat bantu yang sangat handal di dunia pengembangan sistem yang berorientasi objek. Hal ini disebabkan karena UML menyediakan bahasa pemodelan visual yang memungkinkan bagi

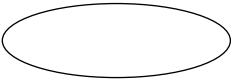
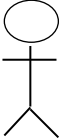

pengembang sistem untuk membuat cetak biru atas visi mereka dalam membentuk yang baku, mudah dimengerti serta dilengkapi dengan mekanisme yang efektif untuk berbagi (*sharing*) dan mengkomunikasikan rancangan mereka dengan yang lain (Munawar ; 2018 : 49).


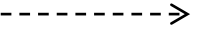
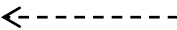
Alat bantu yang digunakan dalam perancangan berorientasi objek berbasis UML adalah sebagai berikut :

1. *Use case* Diagram

Use case diagram merupakan pemodelan untuk kelakuan (*behavior*) sistem informasi yang akan dibuat. *Use case* mendeskripsikan tipikal interaksi antara (pengguna) sebuah *system* dengan sistemnya sendiri melalui sebuah cerita bagaimana sebuah sistem yang dipakai (Munawar ; 2018 : 89).

Tabel II.1. Simbol *Use Case* Diagram

Gambar	Keterangan
	<i>Use Case</i> menggambarkan fungsionalitas yang disediakan sistem sebagai unit-unit yang bertukar pesan antar unit dengan aktor, biasanya dinyatakan dengan menggunakan kata kerja di awal nama <i>Use Case</i> .
	Aktor adalah <i>abstraction</i> dari orang atau sistem yang lain yang mengaktifkan fungsi dari target sistem. Untuk mengidentifikasi aktor, harus ditentukan pembagian tenaga kerja dan tugas-tugas yang berkaitan dengan peran pada konteks target sistem. Orang atau sistem bisa muncul dalam beberapa peran. Perlu dicatat bahwa aktor berinteraksi dengan <i>Use Case</i> , tetapi tidak memiliki <i>control</i> terhadap <i>Use Case</i> .
	Asosiasi antara aktor dan <i>Use Case</i> , digambarkan dengan garis tanpa panah yang mengindikasikan siapa atau apa yang meminta interaksi secara langsung dan bukannya mengindikasikan aliran data.

	Asosiasi antara aktor dan <i>Use Case</i> yang menggunakan panah terbuka untuk mengindikasikan bila aktor berinteraksi secara pasif dengan sistem.
	<i>Include</i> , merupakan di dalam <i>Use Case</i> lain (<i>required</i>) atau pemanggilan <i>Use Case</i> oleh <i>Use Case</i> lain, contohnya adalah pemanggilan sebuah fungsi program.
	<i>Extend</i> , merupakan perluasan dari <i>Use Case</i> lain jika kondisi atau syarat terpenuhi.

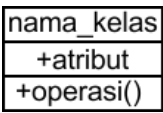
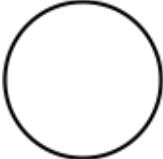
(Sumber : Munawar ; 2018 : 93)





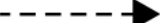
2. *Class Diagram* (Diagram Kelas)

Merupakan hubungan antar kelas dan penjelasan detail tiap-tiap kelas di dalam model desain dari suatu sistem, juga memperlihatkan aturan-aturan dan tanggung jawab entitas yang menentukan perilaku sistem.

Class diagram merupakan diagram statis dari suatu aplikasi. *Class Diagram* tidak hanya digunakan untuk memvisualisasikan, menggambarkan, dan mendokumentasikan berbagai aspek sistem tetapi juga untuk membangun kode eksekusi (*executable code*) dari aplikasi perangkat lunak (Munawar ; 2018 : 101)

Tabel II.2. Simbol *Class Diagram*

No	Simbol	Keterangan
1		Simbol ini adalah simbol untuk sebuah kelas pada struktur sistem. penulisan disana tidak diperbolehkan menggunakan spasi. simbol ini memiliki 3 susunan, yaitu kotak pertama adalah nama kelas, kedua atribut dan terakhir operasi.
2		Lingkaran ini adalah simbol untuk <i>interface</i> atau dalam bahasa indonesianya antar muka. konsep yang digunakan pun sama dengan pemrograman berorientasi object (OOP).



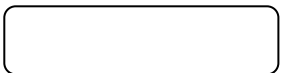
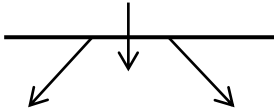
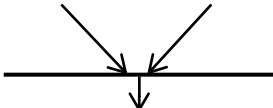
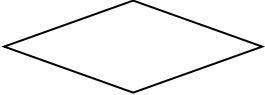
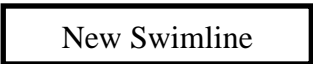
3		Simbol ini sering disebut dengan simbol <i>Association</i> atau dalam bahasa indonesianya yaitu asosiasi. Garis ini adalah garis yang digunakan untuk menghubungkan atau merelasikan kelas satu dengan kelas yang lainnya dengan makna umum.
4		Nama dari simbol ini adalah <i>indirected association</i> atau dalam bahasa indonesianya adalah asosiasi berarah. Simbol ini merupakan simbol relasi antar kelas seperti yang diatas, namun yang membedakan pada relasi ini adalah cara penggunaannya. Simbol ini digunakan jika kelas yang satu digunakan oleh kelas yang lainnya.
5		Simbol ini bernama Generalisasi. Generalisasi digunakan untuk menghubungkan antar kelas dengan arti umum-khusus. Jadi jika ada kelas bermakna umum dan kelas bermakna khusus dapat menggunakan simbol ini.
6		Nama dari simbol ini adalah <i>Aggregation</i> atau dalam bahasa indonesia nya Agregasi. Simbol ini adalah simbol yang menghubungkan antar kelas dengan makna untuk semua bagian. Jadi relasi ini digunakan jika kelas yang satu adalah semua bagian dari kelas yang lainnya.
7		Nama dari simbol ini adalah <i>Dependency</i> atau dalam bahasa indonesia nya ketergantungan. Kadangkala sebuah <i>class</i> menggunakan <i>class</i> yang lain. Umumnya penggunaan <i>dependency</i> digunakan untuk menunjukkan operasi pada suatu <i>class</i> yang menggunakan <i>class</i> yang lain. Sebuah <i>dependency</i> dilambangkan sebagai sebuah panah bertitik-titik.

(Sumber : Munawar ; 2018 : 101)

3. Diagram Aktivitas (*Activity Diagram*)

Activity Diagram bagian penting dari UML yang menggambarkan aspek dinamis dari sistem. Logika prosedural, proses bisnis dan aliran kerja suatu bisnis bisa dengan mudah dideskripsikan dalam *activity diagram* (Munawar ; 2018 : 137).

Tabel II.3. Simbol Diagram Aktivitas

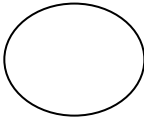
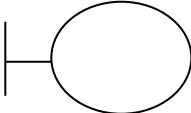
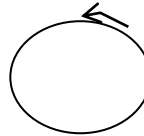

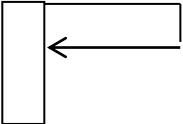


Gambar	Keterangan
	<i>Start point</i> , diletakkan pada pojok kiri atas dan merupakan awal aktifitas.
	<i>End point</i> , akhir aktifitas.
	<i>Activites</i> , menggambarkan suatu proses/kegiatan bisnis.
	<i>Fork</i> (Percabangan), digunakan untuk menunjukkan kegiatan yang dilakukan secara parallel atau untuk menggabungkan dua kegiatan pararel menjadi satu.
	<i>Join</i> (penggabungan) atau rake, digunakan untuk menunjukkan adanya dekomposisi.
	<i>Decision Points</i> , menggambarkan pilihan untuk pengambilan keputusan, <i>true</i> , <i>false</i> .
	Swimlane, pembagian activity diagram untuk menunjukkan siapa melakukan apa.

(Sumber : Munawar ; 2018 : 137)

4. Diagram Urutan (*Sequence Diagram*)

Sequence diagram adalah salah satu *interaction diagram*. Karena *sequence diagram* mengacu kepada obyek, maka sbelum membuat diagram ini class diagram sudah harus teridentifikasi (Munawar ; 2018 : 186).

Tabel II.4. Simbol Diagram Urutan

Gambar	Keterangan
	<i>EntityClass</i> , merupakan bagian dari sistem yang berisi kumpulan kelas berupa entitas-entitas yang membentuk gambaran awal sistem dan menjadi landasan untuk menyusun basis data.
	<i>Boundary Class</i> , berisi kumpulan kelas yang menjadi <i>interface</i> atau interaksi antara satu atau lebih aktor dengan sistem, seperti tampilan <i>formentry</i> dan <i>form</i> cetak.
	<i>Control class</i> , suatu objek yang berisi logika aplikasi yang tidak memiliki tanggung jawab kepada entitas, contohnya adalah kalkulasi dan aturan bisnis yang melibatkan berbagai objek.
	<i>Message</i> , simbol mengirim pesan antar <i>class</i> .
	<i>Recursive</i> , menggambarkan pengiriman pesan yang dikirim untuk dirinya sendiri.
	<i>Activation</i> , mewakili sebuah eksekusi operasi dari objek, panjang kotak ini berbanding lurus dengan durasi aktivitas sebuah operasi.
	<i>Lifeline</i> , garis titik-titik yang terhubung dengan objek, sepanjang <i>lifeline</i> terdapat <i>activation</i> .

(Sumber : Munawar ; 2018 : 186)