

## BAB II

### TINJAUAN PUSTAKA DAN DASAR TEORI

#### 2.1 TINJAUAN PUSTAKA

Dalam penelitian ini, penulis menggunakan beberapa sumber pustaka yang digunakan sebagai pedoman dan pembanding dalam penelitian yang akan penulis lakukan. Pustaka yang digunakan ditinjau dari segi kasus penelitian, objek penelitian, dan metode yang digunakan dalam penelitian.

Penelitian yang pernah dilakukan oleh Bimo Setiawan (2014) menerangkan bahwa sistem prediksi jurusan menggunakan metode *K-Nearest Neighbor* ini menghasilkan output berupa tabel jurusan dan dapat disimpan ke *database* KNN tabel *training* dan dilihat dalam bentuk tabel lengkap.

Ria Agustina Wulandari (2014) menyatakan bahwa sistem prediksi untuk penentuan jurusan bagi mahasiswa baru dengan metode *K-Nearest Neighbor* ini dapat membantu penentuan jurusan di STMIK Akakom. Berdasarkan pengujian data, nilai akurasinya di bawah 80% sehingga mengindikasikan bahwa metode *K-Nearest Neighbor* kurang tepat untuk penentuan jurusan.

Elisabet Yuliana Prasanti (2014) menerangkan dalam implementasi metode *K-Nearest Neighbor* untuk mengklasifikasikan keluarga miskin di Kecamatan Banguntapan mendapatkan hasil bahwa metode *K-Nearest*

*Neighbor* (KNN) cocok digunakan untuk pengklasifikasian keluarga miskin karena tingkat keberhasilan metode KNN lebih dari 80%, dan nilai K yang terbaik adalah  $K = 1$  dengan persentase keberhasilan 92,5%.

Penelitian menggunakan metode *K-Nearest Neighbor* juga pernah dilakukan oleh Ody Andrian (2016) yang menerangkan bahwa aplikasi *data mining* dengan metode *K-Nearest Neighbor* untuk pemilihan laptop dapat memberikan rekomendasi pemilihan laptop secara baik dalam mengimplementasikan metode KNN dari perhitungan manual dan hasil dari aplikasi.

**Tabel 2.1. Perbandingan Hasil Penelitian**

No	Penulis dan Tahun	Objek	Metode	Hasil
1.	Bimo Setiawan (2014)	Prediksi Jurusan	<i>K-Nearest Neighbor</i>	Aplikasi ini menghasilkan output berupa tabel jurusan dan dapat disimpan ke database KNN tabel <i>training</i> dan dilihat dalam bentuk tabel lengkap.
2.	Ria Agustina Wulandari (2014)	Penentuan Jurusan	<i>K-Nearest Neighbor</i>	Sistem ini dapat membantu untuk penentuan jurusan di STMIK Akakom.
3.	Elisabet Yuliana Prasanti (2014)	Data Penduduk	<i>K-Nearest Neighbor</i>	Metode <i>K-Nearest Neighbor</i> (KNN) cocok digunakan untuk pengklasifikasian keluarga miskin karena

No	Penulis dan Tahun	Objek	Metode	Hasil
				tingkat keberhasilan metode KNN lebih dari 80%, dan nilai K yang terbaik adalah $K = 1$ dengan persentase keberhasilan 92,5%.
4.	Ody Andrian (2016)	Laptop	<i>K-Nearest Neighbor</i>	Aplikasi ini dapat memberikan rekomendasi pemilihan laptop secara baik dalam mengimplementasikan metode KNN dari perhitungan manual dan hasil dari aplikasi.
5.	Vicky Prawita (2021)	Pemilihan Jurusan	<i>K-Nearest Neighbor</i>	Aplikasi ini diharapkan mampu membantu siswa SMA yang akan lulus dalam penentuan keputusan pemilihan jurusan kuliah.

Penelitian yang akan penulis lakukan, diharapkan dapat membantu untuk pemilihan jurusan kuliah dan mendapatkan hasil yang sesuai dengan minat siswa.

## 2.2 DASAR TEORI

### 2.2.1 *Data Mining*

*Data Mining* adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam *database*. *Data*

*mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar (Kusrini dan Luthfi, 2009).

Menurut Gartner Group data mining adalah suatu proses menemukan hubungan yang berarti, pola, dan kecenderungan dengan memeriksa dalam sekumpulan besar data yang tersimpan dalam penyimpanan dengan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika (Kusrini dan Luthfi, 2009).

Secara garis besar terdapat dua metoda dalam melaksanakan peran data mining, menurut Kantardzic J.B., M., meliputi (Muflikhah, dkk. 2018):

- Metode Prediksi

Menggunakan beberapa variabel untuk memprediksi nilai yang tidak diketahui atau nilai di masa mendatang dari variabel lain.

- Metode Deskripsi

Menemukan bentuk yang mampu diartikan manusia (*human-interpretable patterns*) yang dapat menjelaskan data tertentu.

Adapun, peran Data Mining, meliputi:

– *Classification [Predictive]*

*Classification* merupakan peran dalam data mining yang menggunakan metode pendekatan prediktif, mempunyai pendefinisian sebagaimana berikut:

- Jika terdapat sekumpulan record (*training set*) dimana setiap record terdiri dari sekumpulan atribut dan satu atribut merupakan kelas.
- Menentukan suatu model untuk atribut kelas sebagai suatu fungsi nilai dari atribut lain.
- Adapun tujuannya adalah record-record yang tidak terlihat sebelumnya ditentukan suatu kelas seakurat mungkin.
- Suatu kumpulan data uji digunakan untuk menentukan keakuratan suatu model. Umumnya, data set yang diberikan dibagi ke dalam sekumpulan data latih dan data uji, dimana data latih digunakan untuk membentuk model dan data uji digunakan untuk mengujinya.

– *Clustering [Descriptive]*

Definisi dari *clustering*, jika diberikan sekumpulan data titik, masing-masing diantaranya memiliki sekumpulan atribut, dan kesamaan ukuran, sehingga dapat ditemukan gugus (*cluster*) sehingga data titik dalam satu *cluster*

(*intracluster*) lebih serupa kepada yang lain. Namun data titik dalam *cluster* yang berbeda (*intercluster*) lebih nampak kurang serupa kepada yang lain.

– *Association Rule Discovery [Descriptive]*

Jika diberikan sekumpulan record dimana masing-masing record terdiri dari sejumlah item dari koleksi yang diberikan. Sehingga perlu dibuat aturan ketergantungan (*dependency rules*) yang akan memprediksi kemunculan item tersebut berdasarkan kebersamaan atas item yang lain.

– *Sequential Pattern Discovery [Descriptive]*

*Sequential pattern discovery* atau penemuan pola secara berurutan, mengandung pengertian sebagaimana berikut, pertama input berupa himpunan objek, dengan setiap objek diasosiasikan dengan waktu kejadiannya. Sedangkan output berupa penemuan rule yang memprediksi kuat ketergantungan berurutan (*sequential dependencies*) diantara kejadian-kejadian yang berbeda  $(A \ B) \ (C) \rightarrow (D \ E)$ . Adapun rule dibentuk oleh pola yang ditemukan. Kebersamaan kejadian dalam pola ditentukan oleh batasan waktu.

– *Regression [Predictive]*

– *Deviation Detection [Predictive]*

### 2.2.2 *K-Nearest Neighbor*

Konsep dasar dari algoritma ini adalah menentukan kelas berdasarkan tingkat kesamaan yang tinggi dari data latih (Muflikhah, dkk. 2018).

Algoritma K-Nearest Neighbors (KNN) adalah algoritma yang berfungsi untuk melakukan klasifikasi suatu data berdasarkan data pembelajaran (*train data sets*), yang diambil dari  $k$  tetangga terdekatnya (*nearest neighbors*). Dengan  $k$  merupakan banyaknya tetangga terdekat. Algoritma KNN adalah salah satu metode yang digunakan untuk analisis klasifikasi, namun beberapa dekade terakhir metode KNN juga digunakan untuk prediksi.

Langkah-langkah untuk menghitung metode *K-Nearest Neighbor* antara lain:

1. Menentukan parameter  $k$
2. Menghitung jarak antara data yang akan dievaluasi dengan semua pelatihan menggunakan rumus *Euclidean Distance* seperti berikut:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

dimana :

- $d(x, y)$  = *dissimilarity*/jarak
- $n$  = dimensi data
- $i$  = variabel data

- $x_i$  = sampel data
  - $y_i$  = data uji atau data testing
3. Mengurutkan jarak yang terbentuk
  4. Menentukan jarak terdekat sampai urutan  $k$
  5. Memasangkan kelas yang bersesuaian
  6. Mencari jumlah kelas dari tetangga yang terdekat dan tetapkan kelas tersebut sebagai kelas data yang akan dievaluasi

### 2.2.3 Pemahaman Minat

Minat adalah salah satu tanda kemantapan dan kesiapan seseorang untuk memilih cita-cita/kariernya dengan adanya dorongan yang kuat dalam belajar, pekerjaan atau tugas-tugas yang dibebankannya. Minat sangat erat sekali hubungannya dengan perasaan suka atau tidak suka, tertarik atau tidak tertarik, senang atau tidak senang.

Minat dapat diukur dengan menggunakan tes dan non tes. Inventory minat merupakan alat untuk mengukur kecenderungan minat seseorang tanpa memberlakukan tes psikologi.

Minat seseorang dapat berpengaruh pada pengambilan keputusan dalam merencanakan masa depan termasuk menentukan program studi yang akan diambil di perguruan tinggi. Terdapat 10 kelompok minat, yaitu:

1. Outdoor

Minat terhadap pekerjaan kegiatan yang biasanya dilakukan di luar rumah atau alam sekitar.

2. Mechanical

Minat terhadap pekerjaan yang berhubungan dengan mesin-mesin atau alat-alat teknik lainnya.

3. Musical

Minat terhadap pekerjaan yang berkaitan dengan musik.

4. Computational

Minat terhadap suatu pekerjaan atau jabatan hitung menghitung.

5. Science

Minat terhadap pekerjaan yang berkaitan dengan ilmu pengetahuan.

6. Persuasive

Minat terhadap pekerjaan yang berhubungan dengan pekerjaan mempengaruhi orang.

7. Artistic

Minat terhadap pekerjaan yang berkaitan dengan kesenian.

8. Literary

Minat terhadap pekerjaan yang berkaitan dengan tulis menulis.

9. Social

Minat terhadap pekerjaan yang berkaitan dengan pelayanan sosial atau membantu orang lain.

10. Clerical

Minat terhadap pekerjaan yang berkaitan dengan pekerjaan administrasi.