

BAB II

TINJAUAN PUSTAKA DAN DASAR TEORI

2.1 Tinjauan Pustaka

Penelitian tentang KNN sendiri sudah banyak dilakukan, baik di bidang kesehatan, agraria, akademik, maupun kesehatan. Fajar (2019) telah melakukan penelitian tentang Prediksi Prestasi Akademik Mahasiswa Berdasarkan Hasil Tes Potensi Akademik dengan algoritma *K-Nearest Neighbor*. Penelitian ini menggunakan data hasil tes potensi akademik pada saat seleksi masuk dan IPK dari semester 1 s.d. 4, dari penelitian ini diperoleh kesimpulan bahwa metode KNN dapat diterapkan untuk memprediksi prestasi akademik mahasiswa berdasarkan hasil tes potensi akademik. Penelitian lainnya dilakukan oleh Slamet (2018) yang meneliti tentang algoritma KNN untuk memprediksi studi mahasiswa, pada penelitian ini diperoleh kesimpulan untuk nilai K diantara 3 s.d 60, diperoleh nilai K terbaik adalah 3, 6, dan 9 agar memperoleh hasil prediksi terbaik. Penelitian lain dilakukan oleh Didik dkk (2016), penelitian ini tentang Perbandingan KNN Dan Naive Bayes Untuk Klasifikasi Tanah Layak Tanam Pohon Jati. Pada penelitian ini disimpulkan bahwa bahwa klasifikasi menggunakan algoritma KNN lebih baik dibandingkan algoritma Naive bayes untuk mengklasifikasi data penanaman pohon jati dari KPH Semarang.

Beberapa penelitian yang sudah pernah dilakukan antara lain:

Tabel 2. 1 Penelitian yang pernah dilakukan

Penulis	Judul	Metode	Parameter
Vincentius Bayu Fajar Kurniawan (2019)	Prediksi Prestasi Akademik Mahasiswa Berdasarkan Hasil Tes Potensi Akademik dengan algoritma K-Nearest Neighbor	K-Nearest Neighbor	Skor tes potensi akademik, ipk sem 1 s.d. 4
Slamet Wiyono dan Taufiq Abidin (2018)	Implementation Of K-Nearest Neighbour (KNN) Algorithm To Predict Student's Performance	K-Nearest Neighbor	IP, IPK, Tempat Tinggal, Jenis Sekolah, Program Studi, Pekerjaan orang tua Status Akademik
Didik Srianto dan Edy Mulyanto (2016)	Perbandingan K-Nearest Neighbor Dan Naive Bayes untuk Klasifikasi Tanah Layak Tanam Pohon Jati	K-Nearest Neighbor, Klasifikasi, Naive Bayes	
Agus Pranoto dkk (2017)	Penerapan Algoritma K-Nearest Neighbors untuk Prediksi keberhasilan studi Mahasiswa Pada Stmik Sinar Nusantara Surakarta	K-Nearest Neighbor	Jenis kelamin, status nikah, status kerja, IP Semester 1-6, capaian SKS, Status Kelulusan.

2.2 Dasar Teori

Dasar teori digunakan untuk memahami definisi, pengertian dasar dan istilah yang digunakan dalam penelitian ini. Berikut dasar teori yang digunakan:

2.2.1 Program Magister

Program Magister adalah jenjang Pendidikan yang ditempuh untuk memperoleh gelar master. Keputusan Senat Akademik UGM Nomor 08/SK/SA/2012 tentang “Kebijakan Akademik Universitas Gadjah Mada” menyebutkan bahwa bagian **Jenis Pendidikan**:

- a. Universitas menyelenggarakan pendidikan akademik meliputi jenjang **sarjana** (strata-1) dan **pascasarjana** (strata-2 dan strata-3).
- b. Universitas dapat menyelenggarakan pendidikan **profesi** berbagai jenjang **spesialisasi**.
- c. Universitas dapat menyelenggarakan pendidikan **vokasi** yang pengelolaannya terpisah dari program pendidikan akademik dan pendidikan profesi.
- d. Universitas dapat bekerjasama dengan organisasi profesi untuk menyelenggarakan proses sertifikasi keahlian tertentu bagi peserta didik.

Beberapa aturan terkait penyelenggaraan program magister tertuang dalam Peraturan Rektor UGM No. 11 Tahun 2016 tentang Pendidikan Pascasarjana.

2.2.2 Penerimaan Mahasiswa Baru

Kebijakan Penerimaan Mahasiswa Baru Program Pascasarjana (S-2 dan S-3) UGM tertuang dalam Peraturan Rektor UGM No. 11 Tahun 2016 tentang Pendidikan Pascasarjana Bab VI Penerimaan Mahasiswa. Sistem Penerimaan Mahasiswa Baru Program Pascasarjana melalui 3 (tiga) jalur, yaitu:

Tabel 2. 2 Jalur Masuk UGM tahun 2016

Jalur	Keterangan
1. Reguler	Diperuntukkan bagi calon pelamar dengan biaya sendiri, calon pelamar dengan biaya dari instansi tempatnya bekerja, dan calon pelamar dari masyarakat umum yang sedang mengajukan beasiswa dari berbagai Kementerian di Republik Indonesia.
2. Kerjasama	Diperuntukkan bagi calon pelamar yang biaya pendidikan dibiayai oleh mitra UGM dan biaya pendidikannya ditanggung oleh institusi tersebut, dibuktikan dengan adanya MoU yang masih berlaku
3. Internasional	Diperuntukkan bagi Warga Negara Asing (WNA)

Pada bagian pertama memuat kriteria penerimaan mahasiswa baru.

Penerimaan mahasiswa baru dilaksanakan melalui sistem seleksi penerimaan mahasiswa baru, dengan mempertimbangkan 3 (tiga) kriteria, yaitu:

- a. Prestasi akademik yang diukur dengan nilai indeks prestasi calon

Mahasiswa pada program Pendidikan sebelumnya;

- b. Potensi akademik yang diukur dengan nilai tes potensi akademik yang masih berlaku;
- c. Kemampuan berbahasa Inggris yang diukur dengan nilai tes bahasa Inggris yang masih berlaku.

Sedangkan pada bagian kedua memuat persyaratan dan prosedur pendaftaran, serta seleksi penerimaan mahasiswa baru program magister. Salah satu syarat diantaranya adalah mempunyai prestasi akademik yang baik ditunjukkan dengan IPK, mempunyai potensi akademik yang masih berlaku, rekomendasi, lama studi, ijin studi, dan kemampuan Bahasa Inggris. Potensi akademik yang masih berlaku, ditunjukkan dengan nilai tes:

- a. Tes Potensi Akademik (TPA);
- b. Tes Potensi Akademik Pascasarjana (PAPs); atau
- c. Tes lain yang setara dan diakui oleh Universitas.

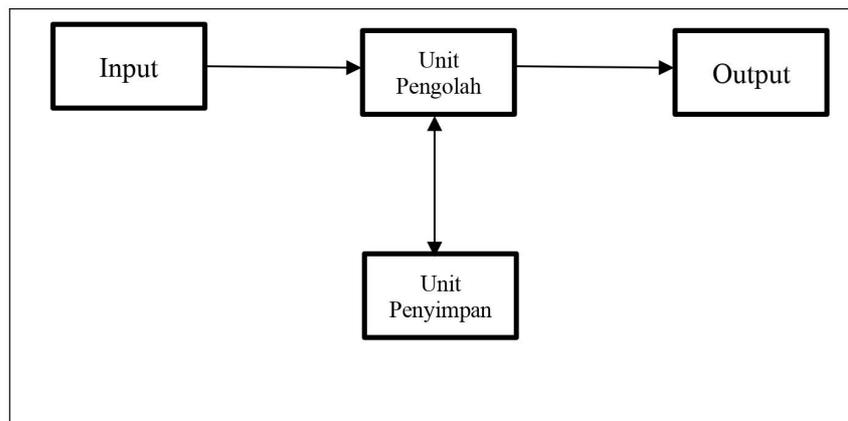
2.2.3 Data dan Informasi

Data dapat didefinisikan sebagai bahan keterangan tentang kejadian-kejadian nyata atau fakta-fakta yang dirumuskan dalam sekelompok lambang tertentu yang tidak acak, yang menunjukkan jumlah, tindakan, atau hal (Sutanta, 2011). Data belum dapat menyampaikan informasi sebelum diolah, misalkan data mahasiswa baru bisa menyampaikan informasi setelah diolah menjadi sebaran peminat mahasiswa baru berdasarkan asal daerah. Untuk proses pengolahan data sendiri diperlukan data.

Informasi merupakan hasil pengolahan data sehingga menjadi bentuk

yang penting bagi penerimanya dan mempunyai kegunaan sebagai dasar dalam pengambilan keputusan yang dapat dirasakan akibatnya secara langsung saat itu juga atau secara tidak langsung pada saat mendatang (Sutanta, 2011). Data dan Pengolah diperlukan untuk menghasilkan informasi. Contoh informasi adalah: Rekapitulasi data mahasiswa aktif dari daerah 3T, rekapitulasi, data lulusan penerima bidikmisi.

Gambar 2.1 ini menunjukkan transformasi data menjadi informasi, input adalah data yang akan diolah oleh unit pengolah yang akan menghasilkan output berupa informasi. Selama proses pengolahan data diperlukan unit penyimpanan (data base) sebagai alat penyimpanan.



Gambar 2. 1 Transformasi data dan informasi

Transformasi data menjadi informasi (Sutanta, 2003)

Pengguna memerlukan informasi, beberapa fungsi informasi antara lain (Sutanta, 2011):

- a. Menambah pengetahuan
- b. Mengurangi ketidakpastian
- c. Mengurangi resiko kegagalan
- d. Mengurangi keanekaragaman/variasi yang tidak diperlukan

- e. Memberi standar, aturan, ukuran, dan keputusan yang menentukan pencapaian sasaran dan tujuan.

Kebutuhan informasi setiap pengguna tidak harus sama, misalkan kebutuhan informasi untuk seorang direktur terkadang berbeda dengan kasubdit informasi dan akademik. Pimpinan memerlukan informasi yang tersaring dan tidak terperinci, misalkan seorang direktur bisa dengan dibuatkan grafik rekap mahasiswa aktif per fakultas. Akan tetapi staff sebaiknya mengerti detail data agar saat ditanyakan detail data oleh pimpinan bisa menjawab detail data tersebut.

2.2.4 Data Mining

Data Mining adalah proses mempekerjakan satu atau lebih teknik pemebelajaran computer (*machine learning*) untuk menganalisis dan mengekstraksi pengetahuan (*knowledge*) secara otomatis (Hermawati, 2013). Hal ini merupakan salah satu langkah dari *Knowledge Discovery in Database* (KDD). Untuk pengambilan keputusan bisa dengan menggunakan data mining yang ada pencarian trend atau pola yang sesuai di database. Sebagai contoh mencari data mahasiswa aktif dari daerah 3T berdasarkan jenis kelamin.

2.2.5 Klasifikasi

Teknik klasifikasi menjadi bagian penting di data mining, yaitu bagaimana mempelajari sekumpulan data sehingga dihasilkan aturan yang bisa mengklasifikasikan atau mengenali data baru yang belum pernah

dipelajari. Klasifikasi adalah proses untuk menyatakan suatu obyek data sebagai salah satu kategori (kelas) yang telah didefinisikan sebelumnya (zaki et al. 2013).

2.2.6 Algoritma KNN

Algoritma KNN merupakan sebuah algoritma pelatihan untuk mengklasifikasikan suatu data baru berdasarkan kategori mayoritas banyaknya K data pelatihan yang terdekat dengannya (*nearest neighbor*). Algoritma ini bertujuan mengklasifikasikan suatu obyek baru berdasarkan sampel pelatihan dan atribut. Dengan memberi titik data baru, kemudian mencari sejumlah K obyek atau titik-titik pelatihan yang terdekat dengan titik baru tersebut. Klasifikasi dilakukan dengan memilih klas terbanyak yang terdapat dalam persekitaran obyek tersebut. Algoritma KNN menggunakan klasifikasi berdasarkan persekitaran obyek sebagai nilai prediksi untuk data baru yang dicari klasnya.

Algoritma KNN bekerja dengan menghitung jarak paling kecil dari data baru dengan semua data latih. Kemudian menentukan klas terbanyak yang muncul berdasarkan nilai K yang telah ditentukan. Dengan menentukan K persekitaran data baru Kemudian dengan menentukan banyaknya K persekitaran data baru tersebut, klas data baru dipilih berdasarkan klas yang paling banyak muncul. Sebagai contoh data pelatihan untuk algoritma KNN yang memuat beberapa atribut *multivariate* dengan variabel F_i yang akan digunakan untuk mengklasifikasikan X. Optimasi parameter dapat menghasilkan nilai K yang. Kasus khusus dimana

klasifikasi diprediksikan berdasarkan training data yang paling dekat (dengan kata lain, $k=1$) disebut algoritma *Nearest Neighbor* (Hermaduanty N & Sri Kusumadewi, 2009).

Nilai k yang terbaik untuk KNN tergantung pada data. Secara umum, nilai k yang tinggi akan mengurangi efek noise pada klasifikasi, tetapi membuat batasan antara setiap klasifikasi menjadi lebih kabur. Nilai k yang bagus dapat dipilih dengan optimasi parameter, misalnya dengan menggunakan *cross-validation*. Kasus khusus dimana klasifikasi diprediksikan berdasarkan data pembelajaran yang paling dekat (dengan kata lain, $k= 1$) disebut algoritma nearest neighbor.

2.2.7 Jarak Euclidian

Nilai kemiripan pada algoritma k -NN dihitung berdasarkan jarak antara data latih dan data uji. Metode *Euclidean distance* merupakan penghitungan jarak pada algoritme k -NN yang paling banyak digunakan oleh para peneliti (Liu & Zhang, 2012).

Rumus *Euclidean Distance*:

$$d_i = \sqrt{\sum_{i=1}^n (p_i - q_i)^2} \dots\dots\dots (2.1)$$

Euclidean Distance terdiri dari empat klasifikasi yaitu :

One -dimensional distance

$$P = (p_x) \text{ dan } Q = (q_x)$$

$$d_i = \sqrt{(p_x - q_x)^2} = |p_x - q_x| \dots\dots\dots (2.2)$$

Two-dimensional distance

$$P = (p_x, p_y) \text{ dan } Q = (q_x, q_y)$$

$$d_i = \sqrt{(p_x - q_x)^2 + (p_y - q_y)^2} \dots\dots\dots (2.3)$$

Three-dimensional distance

$$P = (p_x, p_y, p_z) \text{ dan } Q = (q_x, q_y, q_z)$$

$$d_i = \sqrt{(p_x - q_x)^2 + (p_y - q_y)^2 + (p_z - q_z)^2} \dots\dots\dots (2.4)$$

N-dimensional distance

$$= (p_x, p_y, \dots, p_n) \text{ dan } Q = (q_x, q_y, \dots, q_n)$$

$$d_i = \sqrt{(p_x - q_x)^2 + (p_y - q_y)^2 + \dots + (p_n - q_n)^2} \dots\dots\dots (2.5)$$

(https://en.wikipedia.org/wiki/Euclidean_distance).

Langkah-langkah cara perhitungan KNN berdasarkan *Euclidean Distance*

adalah (Kusumadewi, 2009) :

1. Menentukan parameter k
2. Hitung jarak antara data yang akan dievaluasi dengan semua data pelatihan.
3. Untuk menghitung jaraknya dengan menggunakan rumus *Euclidean*

Distance:

$$d_i = \sqrt{\sum_{i=1}^n (p_i - q_i)^2} \dots\dots\dots (2.6)$$

$$d_i = \sqrt{(p_x - q_x)^2 + (p_y - q_y)^2 + \dots + (p_n - q_n)^2}$$

dimana :

d_i = jarak variabel ke-i

i = variabel data, (i = 1,2,3,...,n)

n = dimensi data

p = data uji, p = (p1,p2,...,pn)

$q = \text{data sampel, } q = (q_1, q_2, \dots, q_n)$

4. Urutkan jarak yang terbentuk (urut naik) dan tentukan jarak terdekat sampai urutan ke-k
5. Pasangkan kategori atau kelas yang bersesuaian.
6. Cari jumlah terbanyak dari tetangga terdekat tersebut, dan tetapkan kategori tersebut sebagai kategori dari data yang dicari.

Proses analisis dan eksplorasi disebut juga dengan *data mining*. Klasifikasi data prediksi keberhasilan studi pada penelitian ini menggunakan algoritma *K-Nearest Neighbor classifier*.

Untuk melakukan evaluasi sistem bisa dilakukan dengan menguji akurasi prediksi sistem. *Confusion matrix* adalah salah satu metode yang dapat digunakan untuk mengevaluasi performa system dalam memprediksi keberhasilan studi. Jika f_{ij} menotasikan jumlah *record* dari kelas i yang berada di kelas j pada saat pengujian, maka pengukuran akurasi (*accuracy*) dapat ditulis dengan persamaan sebagai berikut (Hermawati, 2013):

$$Accuracy = \frac{\text{Jumlah_prediksi_yang_benar}}{\text{Jumlah_prediksi_keseluruhan}} \times 100\% \dots\dots\dots (2.7)$$

Sedangkan tingkat kesalahan (*error rate*) didefinisikan sebagai berikut:

$$Error = \frac{\text{Jumlah_prediksi_yang_salah}}{\text{Jumlah_prediksi_keseluruhan}} \times 100\% \dots\dots\dots (2.8)$$

2.2.8 WEB

World Wide Web atau juga dikenal dengan WEB adalah salah satu layanan yang didapat oleh pemakai komputer yang terhubung ke internet. Secara makna sebuah *website* adalah sekumpulan halaman informasi yang disediakan melalui jalur internet

sehingga bisa diakses diseluruh dunia selama terkoneksi dengan jaringan internet tanpa terbatas ruang dan waktu. *Web* adalah suatu metode untuk menampilkan informasi di internet, baik teks, gambar, suara maupun video yang interaktif dan mempunyai kelebihan untuk menghubungkan (*link*) satu dokumen dengan dokumen yang lainnya (*hypertext*) yang dapat diakses melalui sebuah *browser*. Tahapan-tahapan pengembangan *website* dari awal hingga akhir, adalah *planning*, *design*, *developing/coding*, *upload web*, *promote*, *maintenance web*. Fungsi *website* secara umum adalah fungsi komunikasi, fungsi informasi, fungsi intertainment, dan fungsi transaksi.

2.2.9 MySQL

MySQL adalah suatu *Relational Database Management System* (RDBMS) yang mendukung database terdiri dari sekumpulan relasi atau tabel. MySQL sendiri merupakan aplikasi *open source* yang banyak digunakan untuk penyimpanan data. *Structured Query Language* (SQL) adalah Bahasa yang digunakan untuk berinteraksi dengan *database*.