

BAB 3

METODOLOGI PENELITIAN

3.1. Objek Penelitian

Penelitian ini menjadikan nilai mahasiswa pada mata kuliah dasar umum (MKDU) Bahasa Indonesia di UMKT dari tahun angkatan 2020/2021 hingga 2021/2022 sebagai objek penelitian. Alasan peneliti memilih mata kuliah Bahasa Indonesia karena mata kuliah tersebut mengalami penurunan nilai yang cukup signifikan. Penelitian ini dilakukan di Universitas Muhammadiyah Kalimantan Timur Jl. Ir. Juanda No.15, Sidodadi, Kec. Samarinda Ulu, Kota Samarinda, Kalimantan Timur.

3.2. Metode Pengumpulan Data

Terdapat 2 teknik pengumpulan data yang digunakan dalam penelitian ini, yaitu:

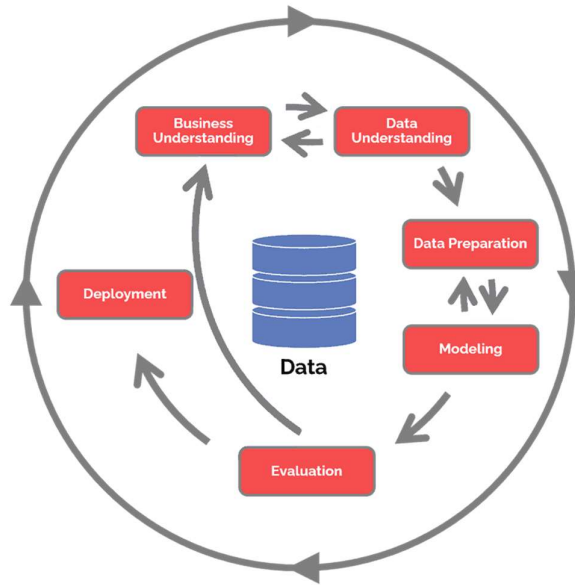
a. Observasi

Teknik ini melakukan pengamatan langsung pada struktur data yang akan diambil dari riwayat perkuliahan di MKDU dan BAA UMKT. Data mahasiswa yang akan diambil dari MKDU dan BAA adalah data mata kuliah Bahasa Indonesia pada tahun angkatan 2020 sampai 2021. Pada tahap ini menghasilkan kumpulan dataset riwayat perkuliahan mahasiswa UMKT pada mata kuliah Bahasa Indonesia dari MKDU dan Unit BAA UMKT.

b. Studi Dokumen

Studi dokumen melakukan pengumpulan data pada dokumen-dokumen yang berkaitan dengan penelitian yang sedang dilaksanakan. Hal ini dilakukan dengan mempelajari jurnal, buku, dan referensi lainnya dalam mendukung penelitian ini.

3.3. Teknik Analisis Data



Gambar 3. 1 Alur Analisis Data

Sumber: (Hotz, 2018)

Pada penelitian ini teknik analisis data yang digunakan adalah CRISP-DM yang memiliki beberapa tahapan yaitu:

1. *Business Understanding* (Pemahaman Bisnis)

Proses klasifikasi nilai mahasiswa pada perkuliahan di UMKT dilakukan dengan teknik *data mining* menggunakan algoritma *Random Forest* dan data yang digunakan adalah data riwayat perkuliahan mahasiswa dari MKDU UMKT dan unit BAA UMKT.

2. *Data Understanding* (Pemahaman Data)

Data yang digunakan diperoleh dari BAA dan MKDU UMKT. Berikut merupakan data yang diperoleh dari MKDU dan BAA yang terlihat dalam tabel dibawah ini.

Tabel 3. 1 Atribut Data MKDU

No	Atribut	Keterangan
1	<i>Profile name</i>	Id mahasiswa pada sistem <i>OpenLearning</i>
2	<i>Learner name</i>	Nama Mahasiswa
3	<i>Learner email</i>	Email Mahasiswa
4	<i>Enrolment ID</i>	Id pendaftaran <i>OpenLearning</i>
5	<i>Institution Membership ID</i>	Id anggota institusi
6	<i>Enrolment Date</i>	Tanggal daftar
7	<i>Completion Date</i>	Tanggal menyelesaikan mata kuliah
8	<i>Time spent on course</i>	Lama waktu mahasiswa berada di mata kuliah
9	<i>Progress</i>	Persentase kemajuan mahasiswa
10	<i>% Course Completed</i>	Persentase kemajuan mahasiswa menyelesaikan mata kuliah
11	<i>Certificate ID</i>	Id sertifikat
12	<i>Comments</i>	Banyaknya komentar mahasiswa selama perkuliahan
13	Tugas 1	Nilai Tugas 1
14	Tugas 2	Nilai Tugas 2
15	Tugas 3	Nilai Tugas 3
16	UTS	Nilai Ujian Tengah Semester

Tabel 3. 2 Atribut Data BAA

No	Atribut	Keterangan
1	NIM	Nomor induk mahasiswa
2	Nama	Nama mahasiswa
3	Jenis Kelamin	Jenis kelamin mahasiswa
4	Nilai Akhir	Nilai akhir mahasiswa
5	Bobot	Bobot nilai akhir
6	Simbol	Simbol nilai akhir

3. *Data Preparation* (Persiapan Data)

Selanjutnya data diidentifikasi, dipilih, dibersihkan, dan diubah ke dalam format yang diinginkan. Data yang digunakan adalah data mahasiswa dari semua program studi UMKT yang mengambil mata kuliah Bahasa Indonesia pada tahun angkatan 2020/2021 dan 2021/2022.

a. *Data Selection* (Seleksi Data) dan *Integration* (Integrasi)

Pada tahap seleksi dan integrasi data ini diperlukan untuk

meningkatkan akurasi dan efisiensi dari algoritma yang digunakan. Kumpulan data yang diseleksi berguna untuk mendapatkan atribut yang cukup dan memiliki pengaruh yang kuat pada data target. Hasil dari seleksi dan integrasi dapat dilihat pada tabel dibawah.

Tabel 3. 3 Atribut Seleksi dan Integrasi Data

No	Atribut	Tipe Data	Keterangan
1	Jenis Kelamin	<i>Numeric</i>	Jenis kelamin mahasiswa
2	<i>Time spent on course</i>	<i>Numeric</i>	Lama waktu mahasiswa berada di mata kuliah
3	<i>Progress</i>	<i>Numeric</i>	Persentase kemajuan mahasiswa dalam mengakses mata kuliah
4	<i>% Course completed</i>	<i>Numeric</i>	Persentase kelas yang diselesaikan mahasiswa
5	<i>Comments</i>	<i>Numeric</i>	Banyaknya komentar mahasiswa selama perkuliahan
6	Tugas 1	<i>Numeric</i>	Nilai tugas 1
7	Tugas 2	<i>Numeric</i>	Nilai tugas 2
8	Tugas 3	<i>Numeric</i>	Nilai tugas 3
9	UTS	<i>Numeric</i>	Nilai ujian tengah semester
10	Simbol	<i>Polynomial</i>	Nilai akhir yang dikonversi menjadi simbol

b. *Data Cleaning* (Pembersihan Data)

Setelah itu, dilakukan pembersihan data dari *noise* yang merupakan kekosongan data, data ekstrim, dan nilai data yang tidak sesuai dengan seharusnya. Hal ini dilakukan untuk menghasilkan data yang bersih dan siap untuk diolah dalam pemodelan.

c. *Data Transformation* (Transformasi Data)

Tahap ini dilakukan untuk merubah format pada kolom data tertentu.

d. *Data Reduction* (Reduksi Data)

Reduksi data dilakukan untuk menyeimbangkan antara data mayoritas dan data minoritas. Pada tahap ini teknik undersampling

digunakan untuk menyeimbangkan jumlah data mayoritas hingga seimbang dengan data minoritas.

4. Modeling (Pemodelan)

- A. Pada tahap ini, langkah pertama adalah membagi data menggunakan *K-Fold Cross Validation* dengan skema *5-fold*, yaitu *data training* dan *data testing*. *Data training* digunakan untuk melatih model, sementara *data testing* digunakan untuk mengevaluasi model yang telah dilatih. Pemodelan yang digunakan dalam penelitian ini adalah algoritma *Random Forest* untuk melakukan prediksi performa akademik mahasiswa di UMKT. Berikut cara kerja dari algoritma *Random Forest Classifier* adalah sebagai berikut (Damanik, 2021):

$$l(y) = \underset{c}{\operatorname{argmax}} \left(\sum_{n=1}^N I_{hn}(y) = c \right)$$

(3.1)

- a. Tahap awal adalah melakukan pengambilan sampel dataset nilai akademik mahasiswa UMKT. Data diambil secara acak (*random*) dari dataset.
- b. Selanjutnya, dataset sampel digunakan untuk membentuk pohon keputusan atau *tree* ke i ($i = 1, 2, \dots, k$). Dalam proses pembangunan *tree*, metode *Classification and Regression Tree (CART)* yaitu *gini impurity* digunakan untuk menentukan setiap *node* pada *tree*.
- c. Langkah 1 dan 2 diulang sebanyak k .
- d. Setelah itu, *data testing* digunakan untuk memprediksi output dari hasil klasifikasi yang telah dilakukan dan menghasilkan sebuah keputusan.
- e. Dilakukan perhitungan suara hasil **voting** untuk setiap target yang akan diprediksi pada setiap **tree**. Pada setiap pohon keputusan, hasil prediksi bisa sama maupun berbeda.

- f. Hasil akhir dari penerapan algoritma *Random Forest Classifier* pada dataset nilai akademik mahasiswa UMKT didapatkan dengan mengambil hasil prediksi target kelas yang paling sering muncul.
- B. Berikut merupakan cara kerja dari seleksi fitur *Analysis of Variance (ANOVA)* adalah sebagai berikut:
- Semua fitur dipilih dari dataset.
 - Fungsi fitur target dari *scikit-learn* dihitung menggunakan *ANOVA F-Score* untuk setiap fitur. Dibawah ini merupakan rumus untuk menghitung *ANOVA*.

$$F = \frac{\text{variance between groups}}{\text{variance within groups}}$$

$$\text{Variance between groups} = \frac{\sum_i^n n_i (\bar{Y}_i - \bar{Y})^2}{(k - 1)}$$

$$\text{Variance within groups} = \frac{\sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2}{(n - k)}$$

(3.2)

- Hasil dari pengujian digunakan untuk melakukan pemilihan fitur yang memungkinkan pembuangan fitur yang tidak terkait dengan variabel target. Fitur yang memiliki pengaruh paling tinggi dengan varian terendah dipilih dalam eksperimen ini dan diuji dengan *SelectKBest()*; *K* mewakili jumlah fitur yang ada untuk dataset akhir.
 - Jumlah fitur dengan peringkat tertinggi digunakan untuk membuat berbagai subset fitur.
5. *Evaluation* (Evaluasi)

Pada tahap evaluasi dilakukan dua kali yaitu sebelum dan sesudah mengimplementasikan seleksi fitur, *confusion matrix* digunakan untuk mengukur akurasi dari hasil pemodelan dengan rumus berikut:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

6. Deployment

Tahap ini akan memberikan rekomendasi dari hasil algoritma *Random Forest* dan hasil seleksi fitur dalam meningkatkan akurasi.

3.4. Peralatan yang Digunakan

Peralatan yang digunakan dalam penelitian ini adalah perangkat keras dan perangkat lunak yaitu:

1. Perangkat Keras (*Hardware*):
 - a. Laptop Dell Inspiron N5110
 - b. Processor Intel Core i7-2670QM
 - c. RAM 6GB DDR3
 - d. Mouse
2. Perangkat Lunak (*Software*):
 - a. Windows 10 Home 64-bit
 - b. Bahasa Pemrograman *Python*
 - c. *Library Sci-kit Learn* (sklearn)

3.5. Jadwal Penelitian

Tabel 3. 4 Jadwal Penelitian

No	Jadwal Penelitian	Bulan/2023					
		Jan	Feb	Mar	Apr	Mei	Jun
Tahap Pra Penelitian							
1.	Menentukan Judul Penelitian						
2.	Menyusun Rumusan Masalah						
3.	Mencari Data Pendukung						
4.	Menyusun Metode Penelitian						
5.	Menyusun Proposal Penelitian						
6.	Seminar Proposal Penelitian						
Tahap Penelitian							
1.	Pengumpulan Data						
2.	Analisis Data						

3.	Pengujian Data <i>Training</i> dan Data <i>Testing</i> dengan Algoritma <i>Random Forest</i> dan Seleksi Fitur <i>Analysis of Variance</i>						
Tahap Akhir Penelitian							
1.	Penyusunan Laporan						
2.	Evaluasi						
3.	Penyempurnaan Laporan						
4.	Seminar Hasil						