

BAB III

METODE PENELITIAN

3.1 Bahan

Bahan yang diperlukan adalah data yang mendukung penelitian ini. Dalam penelitian ini data yang dibutuhkan dan digunakan adalah data faktor penyebab rambut rontok dari *kaggle*.

3.2 Peralatan

Pada prinsipnya setiap desain sistem yang telah dirancang memerlukan sarana pendukung yaitu berupa peralatan yang sangat berperan dalam menunjang penerapan sistem yang didesain terhadap pengolahan data. Komponen yang dibutuhkan antara lain *hardware*, yaitu kebutuhan perangkat keras komputer dalam pengolahan data kemudian *software*, yaitu kebutuhan akan perangkat lunak berupa sistem untuk mengoperasikan sistem yang telah didesain.

a. Perangkat Keras

Adapun spesifikasi minimal perangkat keras (*hardware*) yang digunakan yaitu terdiri dari:

1. Processor Intel Core i3 1.80 Ghz
2. RAM 2GB
3. Harddisk 500GB.

b. Perangkat Lunak

Spesifikasi perangkat lunak (*Software*) yang dibutuhkan adalah:

Microsoft Windows 10 dan Microsoft excel

3.3 Prosedur dan Pengumpulan Data

Pengumpulan data merupakan tahapan persiapan yang harus dilaksanakan terlebih dahulu sebelum dilakukan penelitian. Berikut merupakan aktivitas yang dilaksanakan dalam pengumpulan sumber teori:

1. Studi Literatur

Berfungsi sebagai pendukung dari penelitian yang akan dilaksanakan. Teori-teori yang digunakan bersumber dari buku, jurnal dan penelitian-penelitian sejenis yang dapat mendukung pemecahan masalah dalam penelitian yang dilakukan.

2. Dokumentasi (*Document*)

Merupakan metode pengumpulan data dengan cara membaca, mencatat, mengutip, dan mengumpulkan data-data secara teoritis dari buku-buku dan internet sebagai landasan penyusunan penelitian. Peneliti meminjam buku di perpustakaan, mencari data dari internet juga dilakukan untuk referensi laporan ini, dimana teori tersebut diletakkan pada landasan teori.

3.4 Metode Pendekatan Penyelesaian

KDD (Knowledge Discovery in Database) KDD atau *Knowledge discovery (mining) in databases* Data Mining sering disebut sebagai *Knowledge Discovery in Database (KDD)* . Dalam proses data mining terdapat beberapa teknik untuk menganalisis data seperti Knowledge Discovery in Database (KDD). KDD berhubungan dengan teknik integrasi dan penemuan ilmiah. interpretasi dan dari pola-pola sejumlah kumpulan data. Knowledge discovery in databases (KDD)

adalah keseluruhan proses non-trivial untuk mencari dan mengidentifikasi pola (pattern) dalam data, dimana pola yang ditemukan bersifat sah.

Pada proses Knowledge Discovery Database (KDD) di atas terdapat beberapa tahapan yaitu sebagai berikut:

a) Data Selection

Dalam penelitian ini data yang digunakan adalah 400 data public yang diperoleh dari kaggle. Data terdiri 12 fitur seperti begadang, tingkat tekanan, konsumsi kopi, durasi kerja otak, ujian sekolah, tingkat stres, merk shampo, berenang, mencuci rambut, minyak rambut, ketombe, libido (hormon) di seleksi menjadi 9 fitur yaitu begadang, tingkat tekanan, konsumsi kopi, durasi kerja otak, tingkat stres, berenang, minyak rambut, ketombe, libido (hormon) karena dianggap kurang penting dan memiliki kesamaan dengan fitur lain serta terdiri dari 4 label yaitu few, medium, many, a lot. (Tan, 2019)

stay_up_late	stay_up_late
pressure_level	pressure_level
coffee_consumed	coffee_consumed
brain_working_duration	brain_working_duration
school_assessment	stress_level
stress_level	swimming
shampoo_brand	hair_grease
swimming	Dandruff
hair_washing	Libido
hair_grease	
Dandruff	
Libido	

Tabel 3.1 Seleksi Fitur

b) Pre-processing Cleaning

Sebelum melanjutkan proses data mining data perlu dilakukan pembersihan atau data cleaning pada data selection. Proses data cleaning mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, memperbaiki kesalahan data seperti kesalahan cetak (tipograf). Dimana atribut-atribut terdiri dari begadang, tingkat tekanan, konsumsi kopi, durasi kerja otak, tingkat stres, berenang, minyak rambut, ketombe, libido (hormon).

c) Transformation

Proses mengubah atau menggabungkan data kedalam bentuk yang lebih sesuai untuk proses data mining. Sering kali data yang digunakan dalam proses data mining mempunyai format yang tidak bisa langsung digunakan oleh karena itu diperlukan proses transformasi. (Sutedi, 2020)

3.5 Tahapan Penelitian

1. Identifikasi Masalah

Tujuan Penelitian: tujuan dari penelitian ini adalah untuk mengukur efektifitas metode klasifikasi *Naïve Bayes* dalam memprediksi kemungkinan seseorang mengalami rambut rontok

Rumusan Masalah: apakah metode klasifikasi *Naïve Bayes* efektif digunakan untuk memprediksi rambut rontok ?

2. Studi Literatur

Mengkaji penelitian terdahulu terkait:

- Penyebab rambut rontok (genetik, hormonal, stres, pola makan, dll).
- Penggunaan Naive Bayes dalam klasifikasi data .
- Menentukan variabel-variabel penting untuk dijadikan fitur.

3. Pengumpulan Data

Mengumpulkan data kasus rambut rontok dari dataset public.

fitur yang dikumpulkan:

- a. Begadang
- b. Tingkat Tekanan
- c. Konsumsi Kopi
- d. Durasi Kerja Otak
- e. Tingkat Stres
- f. Berenang
- g. Minyak Rambut
- h. Ketombe
- i. Libido (Hormon)

4. Pra-pemrosesan Data

Pembersihan Data: Menghapus atau menambahkan nilai kosong

Transformasi Data: Konversi data kategorikal menjadi numerik (Encoding).

Seleksi Fitur: Menghapus fitur yang tidak relevan.

5. Pembagian Dataset

Membagi data menjadi:

- Training set (70–80%)
- Testing set (20–30%)

6. Penerapan Algoritma Naive Bayes

Pilih jenis Naive Bayes yang sesuai:

- Gaussian Naive Bayes → jika data numerik dan berdistribusi normal.
- Multinomial/Bernoulli → untuk data kategorikal.

Latih model pada data training.

7. Evaluasi

Menggunakan metrik evaluasi seperti:

- Akurasi
- Precision
- Recall
- F1-score

Evaluasi hasil klasifikasi: misalnya apakah model bisa membedakan antara rambut rontok ringan, sedang, dan parah.

8. Kesimpulan dan Saran

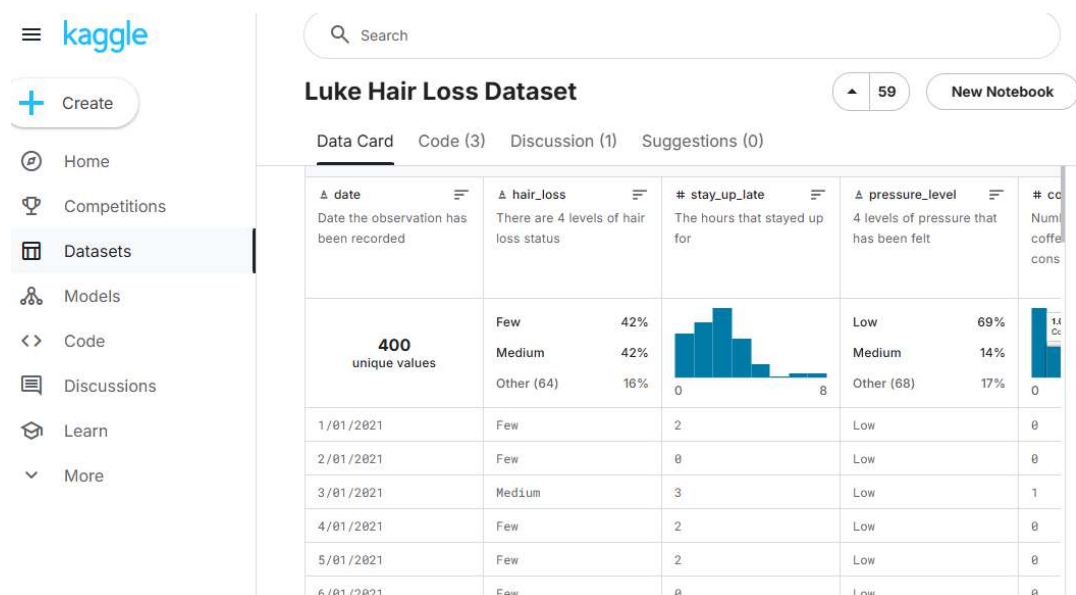
Menyimpulkan efektivitas Naive Bayes dalam klasifikasi rambut rontok.

Memberikan rekomendasi untuk peningkatan model, misalnya:

- Penambahan data
- Eksperimen dengan algoritma lain (SVM, Decision Tree)

3.6 Kaggle

Dataset Kaggle yang terdiri dari 400 data berupa *factor-factor* yang mempengaruhi rambut rontok yang digunakan dalam tahap pelatihan dan pengujian. Gambar 3.1 merupakan contoh *data* yang digunakan.



Gambar 3.1 *Dataset Rambut Rontok*

3.7 Ilustrasi Klasifikasi Menggunakan Naive Bayes

Pada tahap ini akan dilakukan proses klasifikasi dimana akan didapatkan model klasifikasi berdasarkan *input* yang diberikan. Tabel 3.1 merupakan contoh dokumen latih dan uji yang kemudian akan dicari *class* yang paling cocok untuk dokumen uji.

	S8		f _c					
	A	B	C	D	E	F	G	H
1	hair_loss,stay_up_late,pressure_level,coffee_consumed,brain_working_dura							
2	Few,2,Low,0,1,None,Low,Pantene,No,Y,3,None,1							
3	Few,0,Low,0,3,None,Low,Pantene,No,N,1,None,1							
4	Medium,3,Low,1,0,None,Low,Pantene,Yes,Y,2,None,2							
5	Few,2,Low,0,1,None,Low,Pantene,No,N,3,None,3							
6	Few,2,Low,0,1,None,Low,Pantene,No,Y,1,None,2							
7	Few,0,Low,0,3,None,Low,Pantene,Yes,Y,1,None,3							
8	Medium,3,Low,1,1,None,Low,Pantene,Yes,Y,2,None,3							
9	Few,0,Low,0,1,None,Low,Pantene,No,N,3,None,3							
10	Few,2,Low,0,2,None,Low,Pantene,No,N,2,None,4							
11	Medium,1,Low,1,2,None,Low,Pantene,Yes,Y,1,None,5							
12	Medium,0,Low,1,2,None,Low,Pantene,No,N,2,None,5							
13	Medium,2,Low,1,1,None,Low,Pantene,No,N,2,None,5							
14	Medium,0,Low,1,2,None,Low,Pantene,Yes,Y,1,None,5							
15	Medium,2,Low,1,1,None,Low,Pantene,Yes,Y,2,None,5							
16	Medium,3,Low,1,1,None,Low,Pantene,No,N,1,None,4							
17	Few,1,Low,0,1,None,Low,Pantene,No,Y,3,None,4							
18	Few,2,Low,0,0,None,Low,Pantene,Yes,Y,2,None,1							
19	Few,2,Low,0,1,None,Low,Pantene,No,N,1,None,1							
20	Few,2,Low,0,2,None,Low,Pantene,No,N,1,None,1							
21	Few,1,Low,0,0,None,Low,Pantene,Yes,Y,1,None,2							
22	Few,1,Low,0,2,None,Low,Pantene,Yes,Y,2,None,2							
23	Few,0,Low,0,2,None,Low,Pantene,No,N,2,None,3							
24	Few,1,Low,0,1,None,Low,Pantene,No,N,2,None,4							
25	Few,1,Low,0,2,None,Low,Pantene,Yes,Y,2,None,2							

Gambar 3.2 Tabel *dokumen latih* dan *dokumen uji*

Untuk menyelesaikan kasus di atas, tahapan yang dilakukan adalah:

1. Hitung *prior probability* dari dokumen latih. Merujuk pada Tabel 3.1 maka dengan menggunakan rumus pada Persamaan (2.5).
2. Hitung *conditional probability* menggunakan Persamaan (2.6).
3. Hitung *score* dari setiap class pada dokumen uji menggunakan Persamaan (2.2).
4. Prediksi *class* dokumen uji dengan cara memilih class yang memiliki skor terbesar. Dengan menggunakan Persamaan (2.3).

3.8 Timeline Penelitian

NO	KEGIATAN	OKT 2024	NOV 2024	DES 2024	JAN 2025	FEB 2025	MAR 2025	APR 2025	MEI 2025	AGU 2025
1.	Pengajuan Judul									
2.	Studi Literatur									
3.	Pengumpulan Data									
4.	Analisis Data									
5.	Penyusunan Proposal									
6.	Sidang Proposal									
7.	Pengolahan Data									
8.	Sidang Tesis									

Tabel 3.2 Timeline Penelitian