BAB III
METODOLOGI PENELITIAN

3.1 Metode Penelitian

Metodologi penelitian adalah serangkaian prosedur yang dilakukan untuk memperoleh data, menganalisis data, dan mengambil kesimpulan berdasarkan hasil analisis. Berikut adalah penjelasan mengenai metodologi penelitian yang dapat dilakukan untuk penelitian *imputasi missing value* menggunakan dataset MovieLens 100K:

3.1.1 Pengumpulan kebutuhan

Pengumpulan kebutuhan merupakan tahapan dengan melakukan analisis dan perencanaan . Analisa kebutuhan non fungsional adalah sebuah langkah dimana seseorang pembangun perangkat lunak menganalisis sumber daya yang akan menggunakan perangkat lunak yang dibangun. Analisis kebutuhan non fungsional tidak hanya menganalisis siapa saja yang akan menggunakan aplikasi tetapi juga menganalisis perangkat keras dan perangkat lunak agar aplikasi dapat berjalan dengan baik. Analisis non fungsional yang dilakukan dibagi dalam tiga tahapan, yaitu :

(a) Analisis Kebutuhan Pengguna

Aplikasi untuk menentukan konsentrasi skripsi dan rekomendasi bahasa pemrograman ini akan digunakan dengan ketentuan sebagai berikut:

(i) Dapat menggunakan aplikasi yang ada di system operasi Windows.

(ii) Dapat menggunakan media pencarian seperti Mozilla Firefox, Google Chrome, atau browser lain.

(b) Analisis Kebutuhan Perangkat Lunak

Analisis kebutuhan perangkat lunak yang digunakan untuk membangun aplikasi berbasis web adalah sebagai berikut :

(i) OS Windows.

(ii) Browser Internet

(c) Analisis Kebutuhan Perangkat Keras

Analisis kebutuhan perangkat keras yang digunakan untuk membangun sebuah sistem adalah sebagai berikut : Spesifikasi minimum untuk PC :

(i) Processor i5-6500

(ii) Ram 8 GB.

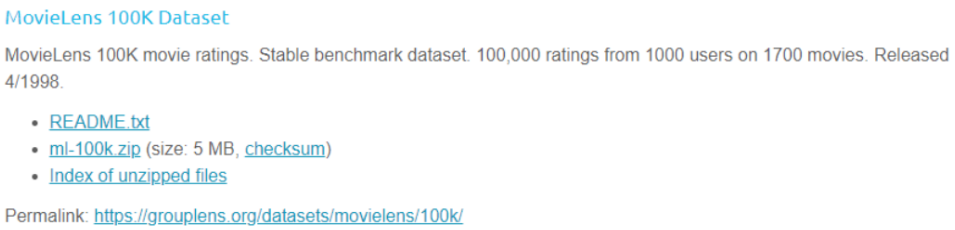
(iii) Hardisk 500 GB.

(iv) Keyboard dan Mouse**.**

Analisa tersebut bukanlah hal yang mutlak, namun merupakan pendapat peneliti tentang minimum penggunaan perangkat keras yang dipakai dalam pengimputasian.

3.1.2 Metode *Data Wrangling*

Dataset yang digunakan dalam penelitian ini adalah Data MovieLens 100K, dan Data Wrangling sebagai metode yang digunakan untuk membersihkan, preprocessing, dan mentransformasi data. dapat diunduh secara gratis dari situs web resmi MovieLens, Seperti yang dilihat di gambar 3.2 . Dataset ini berisi 100.000 peringkat film oleh pengguna MovieLens dan terdiri dari tiga file: file peringkat, file informasi film, dan file pengguna.



Gambar 3.1 *Movielens*

(1) **Cleaning/Preprocessing data**

Setelah data diunduh, langkah selanjutnya adalah membersihkan data. Pada tahap ini, perlu dilakukan identifikasi dan penanganan terhadap data yang hilang atau tidak lengkap, duplikasi data, atau data yang tidak relevan. Misalnya, jika ada baris data yang tidak memiliki nilai atau ada data duplikat, baris data tersebut perlu dihapus.

(2) **Seleksi dan Transforming data**

Setelah data dibersihkan, langkah selanjutnya adalah melakukan seleksi awal transformasi pada data. Pada tahap ini, data yang tidak terstruktur diubah menjadi format yang lebih mudah dianalisis. Contoh dari transformasi data adalah pembuatan kolom baru, mengubah jenis data, dan normalisasi data.

3.1.3 Metode *Systematic sampling*

Systematic sampling atau pengambilan sampel sistematis adalah metode pengambilan sampel acak yang melibatkan pemilihan unit sampel dengan interval tertentu dari populasi target yang diinginkan. Metode ini dilakukan dengan menentukan interval antara unit-unit yang diambil secara acak dari populasi, dan kemudian memilih unit pertama secara acak.

Dalam penelitian ini, jika ingin melakukan systematic sampling pada sebuah populasi sebanyak 1000 user, dengan interval 10, maka setiap 10 user akan diambil sebagai sampel, misalnya user ke-10, ke-20, ke-30, dan seterusnya. Dalam systematic sampling, unit sampel yang pertama diambil secara acak, dan selanjutnya interval ditentukan sesuai dengan jumlah unit sampel yang diinginkan.

3.1.4 Imputasi Data

Setelah tahap sampling data selesai, selanjutnya dilakukan imputasi data. Pada tahap ini, adalah tahap analisis yaitu membandingan hasil imputasi dengan metode satu dengan metode yang lainnya. Berikut

ini adalah contoh pengujian dengan metode imputasi perhitungan manual pada dataset kecil rating film pada tabel 3.1 :

Tabel 3.1 Dataset kecil yang memiliki missing data

ID Penggguna	Umur	Jenis Kelamin	Pekerjaan	Film 1	Film 2	Film 3	Film 4
136	51	Laki - laki	lainnya	4	5	0	5
137	50	Laki - laki	guru	4	4	3	0
138	46	Perempuan	dokter	4	2	2	0
139	20	Laki - laki	siswa	2	3	0	3
140	30	Perempuan	siswi	0	0	0	0

Seperti yang dilihat pada tabel diatas, hasil tidak sempurna dikarenakan masih memiliki data yang hilang. Selanjunya akan diimputasi dengan metode mean dan metode Hot Deck.

(a) Metode Mean

Missing data pada Film 1 diisi dengan rata-rata dari semua score yang di ketahui di film 1 :

$$\frac{4+4+4+2}{4} = 3{,}5$$

Lalu hasil dibulatkan agar mengikuti data yang berisi bilangan bulat menjadi = 4. Hasil imputasi yang didapatkan lalu diisikan ke semua score yang kosong pada fim 1. Pengimputasian dilanjutkan pada score di film 2, 3, 4 dan seterusnya. Berikut adalah data yang sudah diisi dengan menggunakan imputasi konvensional mean :

Tabel 3.2 Dataset kecil yang sudah diimputasi konvensional mean

ID Pengguna	Umur	Jenis Kelamin	Pekerjaan	Film 1	Film 2	Film 3	Film 4
136	51	Laki - laki	lainnya	4	5	3	5
137	50	Laki - laki	guru	4	4	3	4
138	46	Perempuan	dokter	4	2	2	4
139	20	Laki - laki	siswa	2	3	3	3
140	30	Perempuan	siswi	4	4	3	4

Dapat dilihat pada tabel 3.2 bahwa nilai yang diisikan untuk setiap data yang hilang adalah sama. Oleh karena itu semakin banyak presentase data yang hilang pada suatu variable maka akan semakin mengurangi varians dalam suatu variable data.

(b) Metode Hot Deck

Berbeda dengan mean, Hot Deck menggunakan indikator lain sebagai acuan rata-rata. Dan untuk ini saya menggunakan jenis kelamin. Seperti yang dilihat pada film 1, rata rata pengguna perempuan memiliki hasil score 4, maka hasil dari ID pengguna 140 yang memiliki jenis kelamin Perempuan adalah 4. Dan pada Film 3, rata rata pengguna Laki-laki memilki hasil score 3, yang menyebabkan hasil score dari ID pengguna 136 dan 139 adalah 3. Pengimputasian dilanjutkan pada score di film yang belum terisi dan seterusnya. Berikut adalah data yang sudah diisi dengan menggunakan imputasi Hot Deck :

Tabel 3.3 Dataset kecil yang sudah diimputasi Hot Deck

ID Pengguna	Umur	Jenis Kelamin	Pekerjaan	Film 1	Film 2	Film 3	Film 4
136	51	Laki - laki	lainnya	4	5	3	5
137	50	Laki - laki	guru	4	4	3	4
138	46	Perempuan	dokter	4	2	2	1
139	20	Laki - laki	siswa	2	3	3	3
140	30	Perempuan	siswi	4	4	2	1

Perlu diketahui bahwa Hasil imputasi Hot deck tidak cocok pada dataset kecil Dikarenakan Hasil variansi data akan berkurang, Hal ini tidak akan terjadi pada Dataset yang saya gunakan yang berisikan 100,000 data, yang menyebabkanhasil imputansi lebih bervariasi dan kompleks. Selanjutnya adalah membandingkan kedua metode tersebut dengan menggunakan RMSE dan MAE.

3.1.5 Evaluasi

Setelah dilakukan analisis data, langkah terakhir adalah mengambil hasil evaluasi. Kesimpulan ini berdasarkan analisis data dan berisi temuan penelitian atau jawaban terhadap pertanyaan penelitian yang diajukan sebelumnya.

Dalam penelitian ini, metrik evaluasi performa model yang digunakan adalah RMSE *(Root Mean Squared Error)* dan MAE *(Mean Absolute Error).*

1) RMSE

Pada dataset kecil sebelumnya, RMSE memberikan informasi tentang seberapa jauh rata-rata prediksi model dari Dataset awal dalam satuan yang sama dengan nilai dataset yang sudah diimputasi. Semakin rendah nilai RMSE, semakin baik performa model.

$$\text{RMSE} = \sqrt{\frac{(5-5)^2+(0-4)^2+(0-4)^2+(3-3)^2+(0-4)^2}{5}}$$

Melalui perhitungan yang mengikuti persamaan (6) diatas, dapat diketahui bahwa RMSE yang didapat adalah 3,09

2) MAE

Dan selanjutnya jika menghitung dataset kecil tersebut menggunakan MAE, MAE memberikan informasi tentang seberapa jauh rata-rata prediksi model dari dataset awal dalam satuan yang sama dengan nilai MAE. Semakin rendah nilai MAE, semakin baik performa model.

$$\text{MAE} = \frac{(5-5)+(0-4)+(0-4)+(3-3)+(0-4)}{5}$$

Melalui perhitungan yang mengikuti persamaan (7) diatas, dapat diketahui bahwa MAE yang didapat adalah 2,4

3.2 Pembuatan Laporan

Dalam keseluruhan metodologi penelitian, penting untuk memperhatikan validitas dan keandalan data, serta memastikan bahwa analisis dan kesimpulan yang diambil berdasarkan pada metodologi yang tepat.Hasil analisa tersebut akan disusun dalam laporan tugas akhir