

BAB III

METODOLOGI PENELITIAN

3.1 Gambaran Umum Objek Penelitian

Objek dalam penelitian ini adalah data dokumentasi pengujian sistem **internal** yang diperoleh dari pengalaman sebagai Business Analyst Intern di sebuah perusahaan pengembang sistem informasi. Data ini mencakup hasil *System Integration Testing* (SIT), *feedback* pengguna internal, dan dokumentasi visual dari antarmuka pengguna (UI) sebelum dan sesudah dilakukan perubahan desain.

Data tersebut berasal dari beberapa modul dalam sistem internal perusahaan, seperti *Customer Management*, *Lodgement*, *Notification*, *Reporting*, dan modul-modul lainnya yang mengalami proses perbaikan UI selama pengujian. Dokumentasi ini dikumpulkan dalam bentuk laporan naratif, tabel error tracking, serta tangkapan layar (screenshot) UI.

Objek penelitian tidak berfokus pada nama sistem atau organisasi, tetapi pada struktur data dan pola feedback yang muncul dari proses pengujian sistem terhadap UI. Dengan demikian, penelitian ini bersifat generalizable untuk berbagai sistem dengan pendekatan serupa.

Melalui penelitian ini, diupayakan membangun model prediktif kualitas UI berbasis representasi teks dan visual menggunakan model *Transformer* (BERT), sehingga dapat digunakan untuk menilai dan menyarankan perbaikan UI secara otomatis berdasarkan data historis pengujian.

3.1.1 Dataset Penelitian

Dataset yang digunakan pada penelitian ini disusun dari dokumentasi historis pengujian sistem internal yang meliputi hasil *System Integration Testing* (SIT), log interaksi, serta feedback pengguna internal. Dataset ini menjadi dasar pembentukan *data mart* dan pemodelan klasifikasi kualitas UI (bintang 1–5). Rangkuman variabel yang digunakan disajikan pada **Tabel 3.1**.

Tabel 3.1 Tabel Data
Sumber: Dokumentasi Pribadi.

No	Nama Variabel	Deskripsi	Sumber Data	Jenis Data
1	Hasil Uji Sistem	Skor/hasil evaluasi UI dari user testing	Data historis pengujian	Kuantitatif
2	Waktu Interaksi	Lama waktu pengguna menyelesaikan tugas	Log sistem/uji coba	Kuantitatif
3	Jumlah Klik/Error	Banyaknya kesalahan atau interaksi klik	Data uji sistem	Kuantitatif
4	Feedback Pengguna	Respon pengguna terhadap tampilan UI	Kuesioner/Survei	Kualitatif
5	Prediksi Kualitas UI	Hasil prediksi model Transformer	Model penelitian	Kuantitatif

Berdasarkan Tabel Variabel Penelitian, data yang digunakan dalam penelitian ini terdiri dari lima variabel utama yang merepresentasikan hasil pengujian antarmuka pengguna (UI) dari berbagai modul sistem. *Dataset* yang digunakan merupakan data historis hasil *System Integration Testing* (SIT) yang dikompilasi selama proses pengujian internal. Dataset penelitian ini terdiri dari 200 record pengujian, di mana setiap record merepresentasikan satu sesi pengujian UI pada satu modul sistem. Setiap record memuat informasi hasil uji sistem, interaksi pengguna, serta feedback kualitatif yang diberikan setelah pengujian selesai dilakukan.

Data penelitian tidak diperoleh secara langsung dalam jumlah besar, melainkan melalui **proses penggabungan dan pengayaan data** dari beberapa sumber dokumentasi pengujian. Proses penambahan data dilakukan melalui tahapan berikut:

1) **Kompilasi Data Historis Pengujian**

Data awal dikumpulkan dari laporan pengujian sistem (SIT), log sistem, serta dokumen evaluasi UI yang tersedia dalam bentuk file Excel dan laporan internal. Setiap hasil pengujian dicatat sebagai satu entitas pengujian dengan ID unik.

2) **Pengayaan Data Interaksi Pengguna**

Variabel kuantitatif seperti *Waktu Interaksi* dan *Jumlah Klik/Error* diperoleh dari log aktivitas sistem dan catatan hasil uji coba. Data ini mencerminkan tingkat efisiensi dan kompleksitas interaksi pengguna terhadap UI.

3) **Penyesuaian dan Penambahan Record**

Untuk memenuhi kebutuhan analisis dan pelatihan model machine learning, jumlah record diperluas hingga 200 data uji dengan cara:

- a. Menggabungkan hasil pengujian dari beberapa modul sistem.
- b. Menyertakan beberapa skenario pengujian berbeda pada modul yang sama.
- c. Menjaga konsistensi pola distribusi data agar tetap mencerminkan kondisi pengujian nyata.

4) **Integrasi Feedback Pengguna**

Variabel *Feedback Pengguna* diperoleh dari komentar pengguna internal setelah melakukan pengujian UI. Feedback ini bersifat kualitatif dan digunakan sebagai input utama dalam pemodelan berbasis Transformer (BERT).

5) **Pembentukan Variabel Prediksi**

Variabel *Prediksi Kualitas UI* dihasilkan melalui proses pemodelan menggunakan Transformer, yang memanfaatkan kombinasi data numerik dan representasi teks dari feedback pengguna. Variabel ini tidak berasal dari data mentah, melainkan merupakan **output model penelitian**.

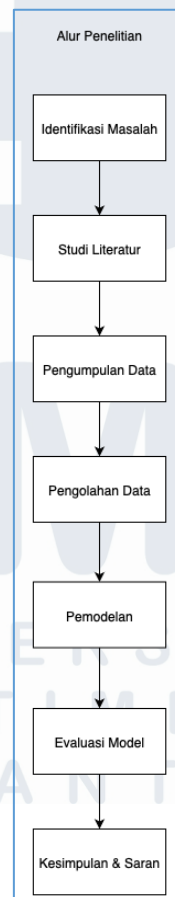
Dengan jumlah **200 record pengujian**, dataset ini dinilai cukup representatif untuk:

- 1) Melatih model prediktif berbasis Transformer.
- 2) Mengidentifikasi pola hubungan antara interaksi pengguna, feedback teks, dan kualitas UI.
- 3) Menghasilkan visualisasi dan rekomendasi perbaikan UI yang berbasis data.

3.2 Metode Penelitian

3.2.1 Alur Penelitian

Penelitian ini dilaksanakan melalui tahapan yang sistematis agar tujuan penelitian dapat tercapai secara terarah. Alur penelitian secara umum terdiri dari:



Gambar 3.1 Alur Penelitian

Sumber: Dokumentasi Pribadi.

1) **Identifikasi Masalah**

Tahap identifikasi masalah diawali dengan pengamatan terhadap proses evaluasi antarmuka pengguna (*User Interface/UI*) pada sistem internal. Evaluasi UI yang dilakukan secara konvensional masih bergantung pada penilaian subjektif evaluator, bersifat manual, dan memerlukan waktu yang relatif lama. Selain itu, data historis hasil pengujian sistem seperti log interaksi, waktu penyelesaian tugas, dan feedback pengguna belum dimanfaatkan secara optimal sebagai dasar evaluasi yang objektif. Permasalahan ini mendorong kebutuhan akan pendekatan berbasis data untuk memprediksi kualitas UI secara lebih konsisten, terukur, dan efisien.

2) **Studi Literatur**

Tahap studi literatur dilakukan untuk memperoleh landasan teoritis dan metodologis yang kuat. Literatur yang dikaji mencakup konsep antarmuka pengguna dan usability, metode evaluasi UI, penerapan machine learning dalam evaluasi sistem, serta penggunaan model *Transformer* khususnya BERT dalam pemrosesan teks. Selain itu, dikaji pula penelitian terdahulu yang memanfaatkan feedback pengguna, data log interaksi, dan visualisasi *dashboard* dalam mendukung pengambilan keputusan. Hasil studi literatur digunakan untuk merumuskan kerangka berpikir penelitian, pemilihan metode, serta justifikasi penggunaan model dan tools yang digunakan.

3) **Pengumpulan Data**

Data penelitian dikumpulkan dari dokumentasi hasil pengujian sistem internal (*System Integration Testing/SIT*). Data yang dikumpulkan meliputi log interaksi pengguna, waktu penyelesaian tugas, jumlah klik atau error, serta *feedback* pengguna terhadap tampilan UI. Setiap data pengujian merepresentasikan satu sesi pengujian pada modul sistem tertentu. Seluruh data historis tersebut kemudian dikompilasi sebagai dataset awal yang menjadi dasar analisis dan pemodelan.

4) **Pengolahan Data**

Tahap pengolahan data mencakup proses pembersihan dan transformasi data agar siap digunakan dalam pemodelan. Proses ini meliputi penghapusan data duplikat dan tidak lengkap, penyamaan format data, serta normalisasi variabel numerik seperti waktu interaksi dan jumlah klik atau error. Selanjutnya, data disusun dalam bentuk data mart menggunakan *Microsoft Excel* untuk mengintegrasikan data numerik dan data teks secara terstruktur. Tahap ini bertujuan memastikan kualitas data dan kesesuaian format input bagi model *machine learning*.

5) **Pemodelan**

Pada tahap pemodelan, dibangun model prediksi kualitas UI menggunakan pendekatan *Transformer*, khususnya BERT, untuk memproses data teks berupa feedback pengguna. Representasi teks hasil *embedding* BERT kemudian digabungkan dengan fitur numerik hasil pengolahan data. Model ini dilatih untuk mempelajari hubungan antara pola interaksi pengguna, *feedback*, dan skor kualitas UI. Hasil dari tahap ini berupa model prediktif yang mampu menghasilkan klasifikasi kualitas UI dalam skala tertentu.

6) **Evaluasi Model**

Evaluasi model dilakukan untuk menilai kinerja dan keandalan model prediksi yang dibangun. Pengukuran performa dilakukan menggunakan metrik evaluasi seperti akurasi, precision, recall, dan F1-score. Selain itu, digunakan confusion matrix untuk melihat distribusi hasil prediksi dan kesalahan klasifikasi antar kategori kualitas UI. Hasil evaluasi ini digunakan untuk menilai sejauh mana model mampu memprediksi kualitas UI secara konsisten dan representatif terhadap data uji.

7) **Kesimpulan dan Saran**

Tahap akhir penelitian adalah penarikan kesimpulan berdasarkan hasil pemodelan dan evaluasi yang telah dilakukan. Kesimpulan mencerminkan pencapaian tujuan penelitian serta kontribusi pendekatan berbasis *Transformer* dalam evaluasi kualitas UI. Selain itu, diberikan saran untuk pengembangan penelitian selanjutnya, seperti penambahan jumlah data, penggunaan model

multimodal yang lebih kompleks, atau penerapan sistem prediksi UI secara langsung dalam proses pengembangan sistem.

3.2.2 Metode Pengembangan Sistem / Metode Data Mining

Dalam penelitian ini, metode yang digunakan disesuaikan dengan kebutuhan penelitian, yaitu:

1) Metode Pengembangan Sistem

Penelitian ini mengadaptasi kerangka kerja **CRISP-DM (Cross-Industry Standard Process for Data Mining)** sebagai metode pengembangan sistem prediksi kualitas antarmuka pengguna. CRISP-DM dipilih karena menyediakan alur kerja yang sistematis, iteratif, dan sesuai untuk penelitian yang melibatkan data historis, pembangunan model prediktif, serta penyajian hasil analisis dalam bentuk visualisasi. Keenam tahapan CRISP-DM diintegrasikan secara langsung ke dalam alur penelitian sebagai berikut.

1) Business Understanding

Tahap business understanding merupakan tahap awal yang bertujuan untuk memahami konteks permasalahan dan tujuan penelitian dari sudut pandang bisnis dan pengguna sistem. Pada penelitian ini, permasalahan utama yang dihadapi adalah proses evaluasi kualitas antarmuka pengguna (UI) yang masih dilakukan secara manual, subjektif, dan memerlukan waktu yang lama. Selain itu, data historis hasil pengujian sistem seperti log interaksi, waktu penyelesaian tugas, dan feedback pengguna belum dimanfaatkan secara optimal.

Pada tahap ini ditetapkan tujuan penelitian, yaitu membangun model prediksi kualitas UI berbasis data historis yang mampu memberikan penilaian secara objektif dan konsisten. Hasil dari tahap ini menjadi dasar dalam menentukan jenis data yang dibutuhkan, pendekatan

pemodelan yang digunakan, serta bentuk keluaran sistem berupa skor kualitas UI dan rekomendasi perbaikan desain.

2) Data Understanding

Tahap data understanding berfokus pada pemahaman terhadap karakteristik data yang digunakan dalam penelitian. Data diperoleh dari dokumentasi hasil pengujian sistem internal (System Integration Testing/SIT) yang mencakup data numerik dan data kualitatif. Data numerik meliputi waktu interaksi pengguna, jumlah klik atau error, serta log aktivitas sistem, sedangkan data kualitatif berupa feedback pengguna terhadap tampilan dan pengalaman penggunaan UI.

Pada tahap ini dilakukan eksplorasi awal data untuk mengetahui jumlah record, distribusi nilai, kelengkapan data, serta potensi permasalahan seperti data hilang atau tidak konsisten. Pemahaman terhadap data ini penting untuk menentukan strategi pengolahan data dan memastikan bahwa dataset yang digunakan relevan dengan tujuan prediksi kualitas UI.

3) Data Preparation

Tahap data preparation bertujuan untuk menyiapkan data agar siap digunakan dalam proses pemodelan. Proses yang dilakukan meliputi pembersihan data, penyamaan format, serta transformasi data. Data yang tidak lengkap atau duplikat dihapus untuk menjaga kualitas dataset. Variabel numerik seperti waktu interaksi dan jumlah klik atau error dinormalisasi ke dalam rentang 0–1 agar memiliki skala yang sebanding dan tidak mendominasi proses pembelajaran model.

Selain itu, data disusun ke dalam bentuk data mart menggunakan *Microsoft Excel* sebagai tahap integrasi awal. Data mart ini menggabungkan seluruh variabel penting dalam satu struktur tabular yang terorganisir, sehingga memudahkan proses analisis dan pemodelan

selanjutnya. Tahap ini menghasilkan dataset akhir yang siap digunakan sebagai input model prediksi.

4) Modeling

Tahap modeling merupakan tahap inti dalam pengembangan sistem prediksi. Pada tahap ini digunakan model Transformer, khususnya BERT, untuk memproses data teks berupa feedback pengguna. Feedback pengguna diubah menjadi representasi numerik melalui proses embedding sehingga dapat dipelajari oleh model. Representasi teks ini kemudian digabungkan dengan fitur numerik hasil pengolahan data sebelumnya.

Model dilatih untuk mempelajari hubungan antara pola interaksi pengguna, feedback teks, dan skor kualitas UI. Proses training dilakukan dengan membagi data menjadi data latih dan data uji untuk memastikan model mampu melakukan generalisasi terhadap data baru. Hasil dari tahap ini adalah model prediktif yang mampu menghasilkan prediksi kualitas UI dalam bentuk skor atau kategori tertentu.

5) Evaluation

Tahap evaluation bertujuan untuk menilai performa model prediksi yang telah dibangun. Evaluasi dilakukan menggunakan metrik kuantitatif seperti akurasi, precision, recall, dan F1-score. Selain itu, confusion matrix digunakan untuk melihat distribusi hasil prediksi dan kesalahan klasifikasi antar kategori kualitas UI.

Hasil evaluasi digunakan untuk menilai apakah model telah memenuhi tujuan penelitian dan mampu memberikan prediksi yang andal. Apabila performa model belum memadai, proses dapat kembali

ke tahap sebelumnya untuk dilakukan penyesuaian, sesuai dengan sifat iteratif dari CRISP-DM.

6) Deployment

Tahap deployment merupakan tahap penerapan hasil model ke dalam konteks yang dapat digunakan oleh pengguna. Pada penelitian ini, hasil prediksi kualitas UI disajikan dalam bentuk dashboard interaktif menggunakan Power BI. Dashboard menampilkan distribusi skor kualitas UI, rata-rata skor per modul, hubungan antara jumlah klik atau error dengan kualitas UI, serta rekomendasi perbaikan desain.

Tahap ini memastikan bahwa hasil penelitian tidak hanya berhenti pada model, tetapi dapat dimanfaatkan secara praktis oleh pemangku kepentingan seperti Business Analyst, QA, dan tim pengembang dalam mendukung pengambilan keputusan desain UI secara data-driven.

2) Metode Data Mining

Pada tahap pemodelan, digunakan algoritma *Transformer* yang berfokus pada kemampuan pemrosesan data sekuensial dan ekstraksi pola dari data historis uji sistem. Model ini dipilih karena terbukti unggul dalam analisis data kompleks dibandingkan metode tradisional.

3.3 Teknik Pengumpulan Data

Teknik pengumpulan data dalam penelitian ini dilakukan dengan beberapa cara agar data yang diperoleh relevan dan dapat mendukung proses pemodelan prediksi kualitas antarmuka pengguna (UI). Adapun teknik pengumpulan data yang digunakan adalah sebagai berikut:

1) Studi Dokumentasi

Mengumpulkan data historis hasil uji sistem yang telah dilakukan sebelumnya. Data ini berupa log interaksi pengguna, waktu penyelesaian tugas,

jumlah kesalahan/klik, serta umpan balik dari pengujian UI. Data dokumentasi digunakan sebagai sumber utama untuk membangun dataset penelitian.

2) Studi Pustaka

Pengumpulan data sekunder dilakukan melalui kajian literatur dari buku, jurnal ilmiah, prosiding konferensi, serta artikel penelitian yang relevan dengan topik UI/UX, machine learning, dan penerapan model Transformer. Studi pustaka ini digunakan untuk memperkuat dasar teori serta membandingkan hasil penelitian dengan penelitian terdahulu.

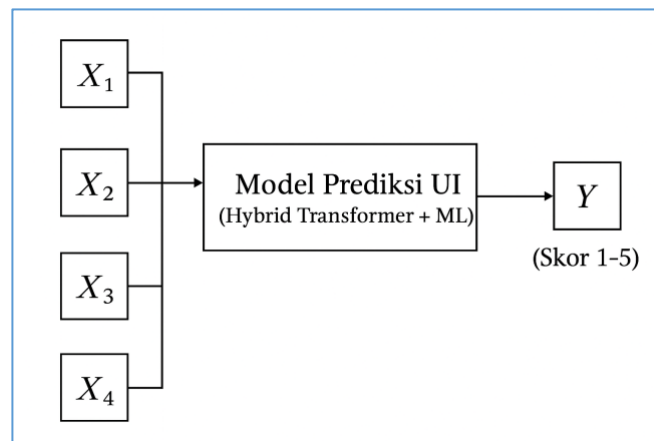
3) Observasi Sistem (jika ada data tambahan)

Apabila diperlukan, peneliti juga dapat melakukan observasi langsung pada sistem yang diuji untuk memperoleh informasi tambahan terkait interaksi pengguna, perilaku penggunaan, dan hambatan yang muncul pada antarmuka.

Dengan teknik pengumpulan data tersebut, penelitian ini diharapkan dapat memperoleh data yang valid dan reliabel untuk proses analisis dan pembangunan model prediksi.

3.4 Variabel Penelitian

Gambar 3.2 menunjukkan bahwa penelitian ini menggunakan empat variabel independen (X_1 – X_4) sebagai masukan (input) yang secara bersama-sama memengaruhi variabel dependen (Y). Seluruh input tersebut diproses oleh Model Prediksi UI (*Hybrid Transformer* + ML), yaitu kombinasi BERT untuk memahami teks feedback dan *machine learning classifier* untuk menghasilkan prediksi skor kualitas UI.



Gambar 3.2 Conceptual Model

Sumber: Dokumentasi Pribadi.

1) Variabel Independen (X)

a) Waktu Interaksi (X1)

X1 merepresentasikan lama waktu yang dibutuhkan pengguna untuk menyelesaikan suatu tugas pada modul tertentu dalam sistem, diukur dalam satuan **detik**. Dalam konteks evaluasi UI, waktu interaksi yang lebih tinggi umumnya mengindikasikan adanya hambatan usability, misalnya navigasi yang tidak jelas, alur kerja terlalu panjang, atau sistem kurang responsif.

Skala data: Rasio, karena memiliki nol absolut dan perbandingan bermakna (misalnya 80 detik dua kali lebih lama dari 40 detik).

b) Jumlah Klik atau Error (X2)

X2 mencatat total interaksi pengguna yang berbentuk klik, termasuk kesalahan klik (mis-click) maupun error yang muncul saat pengguna menjalankan tugas. Nilai X2 yang tinggi biasanya menjadi indikator bahwa UI kurang efisien atau membingungkan, karena pengguna perlu mencoba berulang kali atau melakukan koreksi akibat kesalahan.

Skala data: Rasio, karena berupa hitungan dengan nol absolut dan dapat dibandingkan secara proporsional.

c) Log Aktivitas Sistem (X3)

X3 merupakan jumlah aktivitas atau event yang tercatat pada sistem saat pengguna berinteraksi, misalnya membuka halaman, submit form, membuka menu, validasi error, dan event lain yang terekam. Variabel ini membantu menggambarkan kompleksitas proses interaksi yang terjadi selama pengujian. Semakin tinggi log aktivitas, bisa berarti prosesnya lebih kompleks, banyak percobaan langkah, atau terdapat hambatan yang memicu aktivitas berulang.

Skala data: Interval (sesuai rancangan penelitianmu), karena digunakan sebagai indikator kuantitatif yang selisihnya bermakna untuk membandingkan tingkat aktivitas antar pengujian.

d) Feedback Pengguna (X4)

X4 berisi komentar atau umpan balik kualitatif dari pengguna setelah melakukan pengujian modul UI, misalnya terkait kejelasan tampilan, konsistensi tombol, kemudahan navigasi, atau performa sistem (lambat atau responsif). Pada diagram, X4 menjadi input penting karena diproses dengan **Transformer (BERT)** untuk menangkap makna dan konteks bahasa dalam feedback.

Secara konseptual, feedback dapat dinyatakan dalam tingkat kepuasan (contoh: sangat tidak puas sampai sangat puas) sehingga dipetakan sebagai **skala ordinal**.

Skala data: Ordinal, karena mengandung urutan tingkat penilaian.

2) Proses pada Model Prediksi UI (Hybrid Transformer + ML)

Sesuai diagram, keempat variabel X1–X4 tidak berdiri sendiri, tetapi digabung sebagai satu paket input ke model:

- 1) **X4 (Feedback Pengguna)** diproses oleh **BERT** untuk menghasilkan representasi numerik (embedding) yang menangkap konteks kata.

- 2) **X1, X2, X3 (numerik/log)** diproses melalui tahap persiapan data (misalnya normalisasi) agar skala fitur sebanding.
- 3) Hasil dari fitur teks (embedding BERT) dan fitur numerik kemudian digabung (hybrid features) dan diteruskan ke algoritma klasifikasi machine learning untuk menghasilkan label prediksi.

3) Variabel Dependen (Y): Kualitas Desain UI (Skor 1–5)

Output akhir pada diagram adalah **Y**, yaitu skor kualitas desain UI dalam bentuk klasifikasi **1 sampai 5**. Skor ini merepresentasikan tingkat kualitas UI berdasarkan pola yang dipelajari dari gabungan data historis (X1–X3) dan persepsi pengguna (X4).

Makna skornya:

1. **Skor 1 (Sangat Buruk):** UI sulit digunakan, banyak error, navigasi membingungkan, butuh perbaikan besar.
2. **Skor 2 (Buruk):** masih ada masalah signifikan seperti alur kurang jelas, respons lambat, atau inkonsistensi elemen.
3. **Skor 3 (Cukup):** dapat digunakan, tetapi masih ada hambatan usability; perlu perbaikan minor hingga menengah.
4. **Skor 4 (Baik):** sebagian besar elemen jelas, responsif, konsisten; hanya butuh penyempurnaan kecil.
5. **Skor 5 (Sangat Baik):** UI sangat efisien, stabil, jelas, dan nyaman digunakan tanpa masalah berarti.

Skala data: Ordinal, karena skor menunjukkan urutan kualitas (1 lebih rendah dari 2, dan seterusnya), tetapi jarak antar skor tidak selalu sama secara matematis.

1) **Hipotesis Penelitian**

Berdasarkan variabel yang digunakan, hipotesis penelitian dapat dirumuskan sebagai berikut:

Landasan teori dan penelitian terdahulu (Zhang et al., 2020; Xie et al., 2021; Wang & Fang, 2022; Nguyen et al., 2021; Liu et al., 2019), model berbasis *deep learning*, khususnya arsitektur Transformer seperti BERT terbukti mampu menghasilkan akurasi yang lebih tinggi dalam tugas prediksi berbasis teks dan multimodal dibandingkan model tradisional seperti Random Forest dan Logistic Regression.

3.5 Teknik Analisis Data

Teknik analisis data digunakan untuk mengolah, menguji, dan menarik kesimpulan dari data historis uji sistem yang telah dikumpulkan. Analisis dilakukan agar hasil penelitian dapat menjawab rumusan masalah dan menguji hipotesis yang telah ditetapkan.

3.5.1 Analisis Data Historis Uji Sistem

Pada tahap ini dilakukan pengolahan data mentah hasil uji sistem yang meliputi waktu interaksi, jumlah klik/error, log aktivitas sistem, dan feedback pengguna. Langkah-langkah analisis meliputi:

- 1) **Pembersihan Data (Data Cleaning):** Menghapus data duplikat, data tidak lengkap, dan data tidak relevan.
- 2) **Transformasi Data:** Melakukan normalisasi atau standarisasi agar variabel memiliki skala yang sebanding.
- 3) **Pembuatan Data Mart:** Mengorganisir data historis ke dalam format yang terstruktur menggunakan Microsoft Excel agar mudah dianalisis lebih lanjut.
- 4) **Eksplorasi Data (EDA):** Melakukan analisis deskriptif untuk memahami pola awal, distribusi data, dan korelasi antarvariabel.

3.5.2 Analisis Prediktif dengan Model Transformer

Tahap ini fokus pada pemodelan prediksi kualitas UI menggunakan algoritma Transformer. Adapun langkah-langkah yang dilakukan adalah:

- 1) **Pembangunan Model:** Melatih model Transformer dengan dataset hasil pengolahan.
- 2) **Validasi Model:** Membagi dataset menjadi *training set* dan *testing set* untuk menguji kemampuan generalisasi model.
- 3) **Evaluasi Model:** Menggunakan metrik evaluasi seperti akurasi, precision, recall, dan F1-score untuk mengukur performa model.
- 4) **Interpretasi Hasil:** Membandingkan hasil prediksi dengan penilaian aktual (ground truth) untuk menilai efektivitas pendekatan.

Hasil pemodelan Transformer diintegrasikan ke dalam simulasi antarmuka berbasis web/aplikasi. Pada tahap ini, sistem memungkinkan pengguna untuk **mengunggah screenshot UI** dari modul tertentu. Model kemudian akan menganalisis representasi visual tersebut bersama dengan data historis terkait (log, feedback, klik/error), dan menghasilkan **prediksi tingkat kualitas UI**.

Prediksi ditampilkan dalam dua bentuk:

1. **Klasifikasi 5 kelas (bintang 1–5)** yang mewakili mutu UI dari sangat buruk hingga sangat baik.
2. **Keterangan user friendly atau tidak user friendly**, yang diperoleh dengan mengelompokkan hasil klasifikasi (contoh: bintang 1–2 = tidak user friendly, bintang 3 = cukup, bintang 4–5 = user friendly).

Visualisasi hasil prediksi dapat ditampilkan langsung pada dashboard interaktif di Tableau/Power BI dalam bentuk rating bintang, distribusi kelas, dan perbandingan antar modul UI.

Dengan teknik analisis ini, penelitian diharapkan dapat memberikan gambaran objektif mengenai kemampuan model Transformer dalam memprediksi kualitas desain antarmuka pengguna.