

## BAB III

### METODOLOGI PENELITIAN

#### 3.1 Gambaran Umum Objek Penelitian

Penelitian ini dilakukan untuk menjawab kebutuhan dalam menganalisis pola perilaku tamu serta memahami faktor-faktor yang memengaruhi lama menginap di Hotel XYZ. Hotel ini merupakan salah satu properti perhotelan yang dikelola oleh PT Indonesia Dunia Berkreatif, sebuah perusahaan teknologi yang berfokus pada pengembangan perangkat lunak manajemen hotel dan solusi digital untuk industri perhotelan. Pendekatan *Machine Learning* diterapkan sebagai metode analisis prediktif yang relevan dan sesuai untuk menghasilkan model yang mampu memprediksi durasi menginap tamu secara akurat dengan memanfaatkan data historis tahun 2024. Hasil penelitian diharapkan dapat memberikan rekomendasi berupa *business insight* yang signifikan, terutama dalam memproyeksikan perilaku tamu pada tahun 2026.

##### 3.1.1 Sejarah Perusahaan

PT Indonesia Dunia Berkreatif merupakan perusahaan yang berfokus pada pengembangan teknologi perangkat lunak untuk mendukung operasional industri perhotelan di Indonesia. Perusahaan ini hadir sebagai bagian dari upaya mendorong inovasi digital di sektor perhotelan, dengan menyediakan sistem manajemen yang efisien, terintegrasi, dan sesuai dengan kebutuhan industri modern.

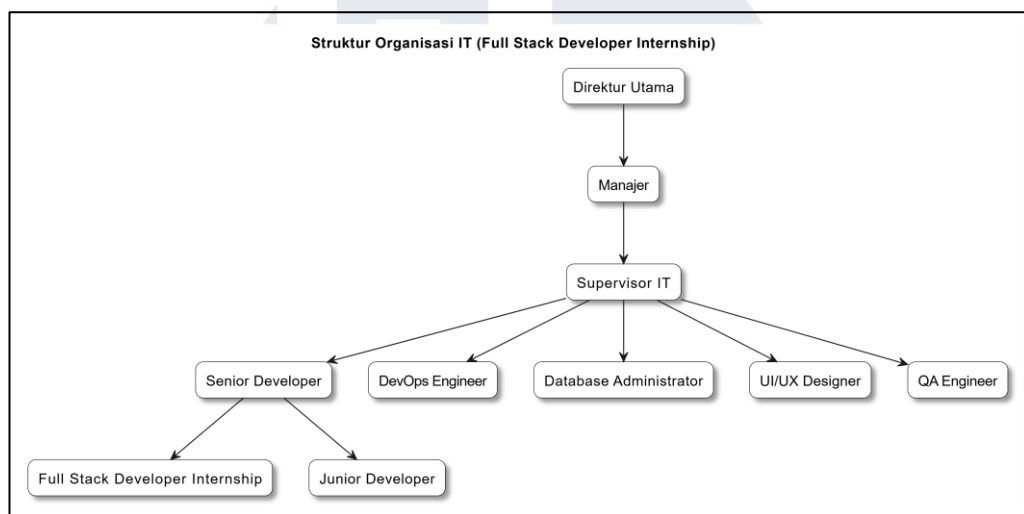


Gambar 3. 1 Logo Indonesia Dunia Berkreatif

Perusahaan memiliki visi untuk berkontribusi nyata dalam pengembangan industri perhotelan di Indonesia dan kawasan Asia Pasifik melalui sistem manajemen berbasis *cloud* yang inovatif dan berkelanjutan.

Dalam menjalankan visinya, perusahaan berkomitmen menghasilkan perangkat lunak yang selalu relevan dengan perkembangan teknologi global serta menyediakan layanan pendampingan yang profesional dan responsif. Perusahaan ini berfokus membantu mitra perhotelan meningkatkan efisiensi operasional, pengelolaan sumber daya, dan kinerja melalui otomatisasi sistem, sekaligus memastikan pemanfaatan teknologi yang optimal agar klien dapat memperoleh manfaat maksimal dari solusi digital yang diterapkan.

### 3.1.2 Struktur Organisasi Perusahaan



Gambar 3. 2 Struktur Organisasi Perusahaan

Gambar 3.2 di atas menjelaskan struktur organisasi divisi IT pada perusahaan yang menjadi bagian penting dalam pelaksanaan penelitian. Struktur ini menunjukkan hubungan hierarki yang jelas antara manajemen dan tim teknis, di mana setiap bagian memiliki peran terkoordinasi untuk mendukung pengembangan sistem berbasis teknologi. Divisi IT berfungsi sebagai pusat *monitoring* dan pengelolaan terhadap seluruh klien hotel yang menggunakan aplikasi dan perangkat lunak yang dikembangkan oleh perusahaan, sehingga memastikan setiap sistem berjalan stabil, terintegrasi, dan sesuai dengan kebutuhan operasional masing-masing hotel, termasuk dalam mendukung penelitian yang berfokus pada penerapan metode *Machine Learning* untuk analisis data dan optimasi sistem. Melalui struktur organisasi yang teratur ini, kegiatan penelitian dapat berjalan secara sistematis, terarah,

dan kolaboratif karena adanya dukungan koordinasi yang baik antarbagian dalam tim IT.

### 3.2 Metodologi Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan menggunakan data internal yang dimiliki oleh Hotel XYZ sebagai sumber utama dalam proses analisis. Data yang digunakan merupakan laporan manajemen operasional hotel, yang mencakup informasi mengenai pola reservasi kamar, rata-rata durasi menginap tamu, serta sistem manajemen data yang diterapkan dalam kegiatan operasional sehari-hari. Seluruh data diperoleh secara langsung dari pihak manajemen Hotel XYZ melalui proses koordinasi dan verifikasi untuk memastikan keakuratan serta kelengkapan informasi yang akan digunakan dalam penelitian.

Proses pengumpulan data dilakukan secara sistematis agar informasi yang diperoleh dapat dipertanggungjawabkan dan relevan dengan tujuan penelitian. Data mentah yang diperoleh dari laporan manajemen hotel kemudian melalui tahap pengolahan dan penyesuaian untuk memastikan keseragaman format serta konsistensi antaratribut data. Tahapan ini melibatkan proses pembersihan data (*data cleaning*) untuk menghilangkan entri yang tidak lengkap atau tidak valid, serta tahapan transformasi data agar dapat digunakan dalam proses analisis dan pemodelan prediksi.

Selain pengumpulan data, penelitian ini juga didukung oleh studi literatur yang komprehensif. Peneliti melakukan telaah terhadap berbagai referensi ilmiah baik dari jurnal nasional maupun internasional, buku teks terbaru, serta hasil penelitian terdahulu yang relevan dengan bidang *data mining* dan *Machine Learning*, khususnya yang menerapkan algoritma *Random Forest* dan *Gradient Boosting* pada konteks analisis perilaku pelanggan di industri perhotelan dan pariwisata. Studi pustaka dilakukan melalui berbagai *database* ilmiah dan portal terbuka agar referensi yang digunakan bersifat valid, mutakhir, serta selaras dengan kerangka penelitian.

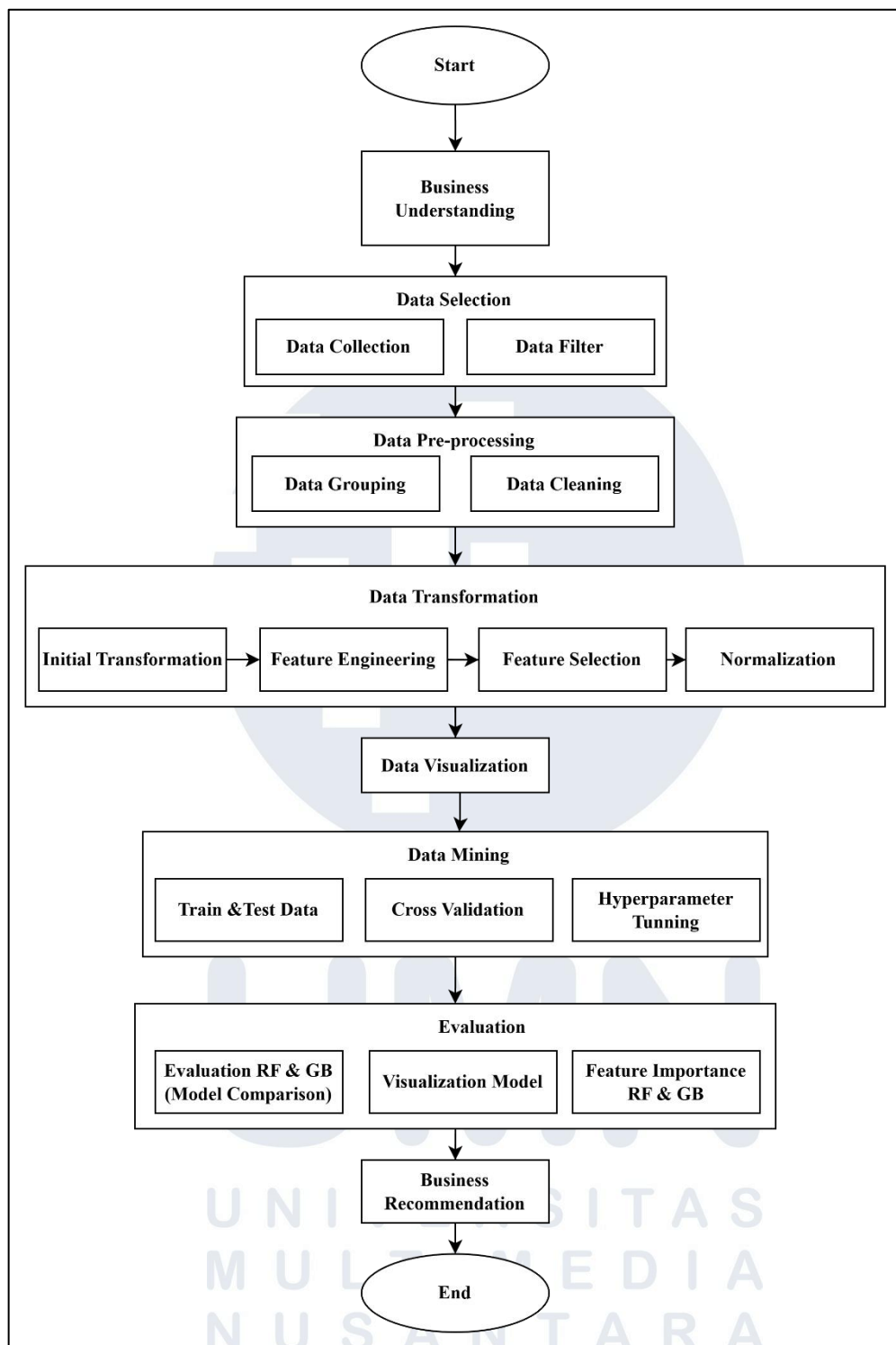
Dalam proses perancangan dan analisis, penelitian ini menerapkan kerangka kerja *Knowledge Discovery in Database (KDD)* yang terdiri dari tahapan seleksi

data, *preprocessing*, transformasi, dan *data mining*. Setiap tahapan dilakukan secara sistematis untuk menghasilkan model prediksi lama menginap tamu yang akurat. Pada tahap *data mining*, digunakan pendekatan *supervised learning* dengan algoritma *Random Forest* dan *Gradient Boosting* untuk membangun model prediksi. Selain itu, pada tahap *feature engineering*, atribut kategorikal diubah menjadi bentuk numerik melalui proses *encoding* agar dapat diproses oleh algoritma yang digunakan.

Evaluasi model dilakukan menggunakan metrik statistik seperti *Root Mean Square Error* (RMSE), *Mean Absolute Error* (MAE), dan koefisien determinasi ( $R^2$ ) guna menilai tingkat akurasi dan reliabilitas hasil prediksi. Validasi model dilakukan melalui teknik *cross-validation* untuk memastikan performa model stabil terhadap variasi data. Seluruh tahapan penelitian ini dirancang secara sistematis dan objektif dengan tujuan menghasilkan model prediksi yang tidak hanya akurat secara matematis, tetapi juga bermanfaat secara praktis bagi manajemen Hotel XYZ dalam pengambilan keputusan strategis terkait pengelolaan reservasi dan peningkatan efisiensi operasional.

### 3.3 Alur Penelitian

Alur penelitian dimulai dengan tahapan awal penelitian yang telah disesuaikan dengan metode *Knowledge Discovery in Database* (KDD), seperti ditunjukkan pada gambar dibawah berikut ini.



Gambar 3. 3 Alur Penelitian

Gambar diatas merupakan alur penelitian yang mengikuti lima tahapan utama KDD. Setiap tahapan dirancang secara iteratif untuk memastikan kualitas

data dan model yang dihasilkan dapat memberikan *insights* yang *actionable* bagi manajemen Hotel XYZ.

### 3.3.1 Business Understanding

Tahap pertama adalah *business understanding*, yang merupakan fondasi dari seluruh proses penelitian. Tahap ini bertujuan untuk memahami kebutuhan bisnis Hotel XYZ dan merumuskan tujuan penelitian secara jelas. Permasalahan utama Hotel XYZ adalah belum tersedianya sistem prediksi lama menginap tamu (*length of stay*) yang akurat, sehingga berdampak pada ketidakefisienan operasional seperti perencanaan *housekeeping*, penjadwalan staf, dan strategi harga kamar. Pada tahap ini ditetapkan tujuan bisnis, yaitu mengembangkan model prediksi LoS berbasis *Machine Learning* dengan pendekatan *Random Forest* dan *Gradient Boosting* untuk membantu hotel mengoptimalkan manajemen okupansi serta meningkatkan profitabilitas melalui pengambilan keputusan berbasis data.

### 3.3.2 Data Selection

Tahapan ini mencakup proses data *collection* dan data *filtering* terhadap data transaksi internal Hotel XYZ yang disimpan dalam basis data MariaDB dan dikelola melalui *phpMyAdmin*. Proses data *collection* dilakukan dengan mengekstrak variabel-variabel penting seperti *room night*, *number of guests*, *room revenue*, *food and beverage revenue*, dan *Hotel revenue* menggunakan *Structured Query Language* (SQL) terstruktur. Data ini diekstraksi ke dalam bentuk dokumen *Comma-Separated Values* (CSV) dan *load* ke dalam aplikasi *Jupyter Notebook* yang menggunakan bahasa pemrograman *python*. Setelah itu dilakukan data *filtering* untuk memilih data transaksi selama periode tahun 2024 agar analisis bersifat konsisten secara temporal dan representatif terhadap kondisi terkini hotel. Pada tahap ini juga dilakukan pelabelan target (*data labeling*) dengan menetapkan kolom lama menginap tamu sebagai variabel dependen yang merepresentasikan lama menginap tamu. Hasil seleksi ini kemudian menjadi dataset utama yang akan diolah pada tahap selanjutnya.

### 3.3.3 Data Preprocessing

Tahap data *preprocessing* merupakan tahapan penting yang bertujuan untuk meningkatkan kualitas data sebelum dilakukan analisis prediktif. Pada tahap ini, terlebih dahulu dilakukan proses data *grouping* untuk menyederhanakan variabel segmentasi tamu dengan menggabungkan berbagai kategori *description* ke dalam beberapa kelompok segmen utama yang representatif, sehingga mengurangi kompleksitas data dan mempermudah proses analisis. Selanjutnya, dilakukan proses data *cleaning* untuk mengidentifikasi dan menangani *missing values* menggunakan teknik imputasi, serta penghapusan *duplicate entries* dan anomali data yang tidak konsisten guna meningkatkan reliabilitas *dataset*. Berdasarkan implementasi menggunakan *python*, proses data *cleaning* mencakup penggantian nilai tak terhingga (*infinite*) menjadi NaN, penghapusan baris data yang memiliki nilai kosong atau nol pada kolom target lama menginap tamu, serta penghapusan data anomali dengan nilai negatif. Selain itu, dilakukan imputasi menggunakan nilai median untuk kolom numerik yang memiliki nilai hilang dan pengisian nilai “*unknown*” untuk kolom kategorikal.

### 3.3.4 Data Transformation

Tahap transformasi dilakukan melalui empat komponen yang saling terkait. Pertama, *initial transformation*, yaitu penurunan fitur waktu dari data tanggal hotel berupa tahun, bulan, tanggal, indeks hari dalam pekan, serta indikator akhir pekan untuk menangkap pola musiman. Kedua, *feature engineering*, yaitu pembentukan variabel turunan yang relevan secara bisnis seperti *revenue per booking*, rata-rata pendapatan per tamu, dan rasio pendapatan makanan dan minuman terhadap total pendapatan, serta pembuatan variabel kuartal dan *dummy* musiman. Ketiga, *feature selection*, yaitu pemilihan fitur inti dan fitur turunan yang paling relevan dengan prediksi lama menginap berdasarkan pertimbangan *domain* dan korelasi awal. Keempat, normalisasi, yaitu skala *Min–Max* pada fitur numerik agar rentang nilai seragam dan tidak menimbulkan dominasi fitur bernilai besar. Setelah transformasi, dilakukan *exploratory data analysis* untuk memahami distribusi dan kecenderungan data.



Visualisasi meliputi ringkasan pola musiman per bulan, hubungan antara jumlah tamu dan pendapatan hotel, distribusi komposisi deskripsi bila tersedia, serta peta korelasi antarvariabel untuk menilai kedekatan hubungan dengan target.

### 3.3.5 Data Mining

Tahap selanjutnya adalah *data mining*, yang merupakan inti dari proses KDD dimana algoritma *Machine Learning* diterapkan untuk membangun model prediktif. Penelitian ini mengimplementasikan *Random Forest* serta *Gradient Boosting*. Tahap pemodelan dimulai dengan pembagian data menjadi data latih dan data uji dengan proporsi delapan puluh banding dua puluh agar evaluasi kinerja model bersifat objektif. Validasi silang menggunakan lima lipatan diterapkan pada data latih untuk menilai konsistensi performa model pada berbagai pembagian data. Selanjutnya dilakukan penalaan hiperparameter menggunakan pencarian *grid* pada kedua algoritma, yaitu *Random Forest* dan *Gradient Boosting*, untuk memperoleh kombinasi parameter terbaik seperti jumlah pohon, kedalaman maksimum, tingkat pembelajaran, dan *subsampling*. Tujuan tahap ini adalah meminimalkan kesalahan, meningkatkan akurasi, serta mengendalikan risiko *overfitting* sehingga model memiliki kemampuan generalisasi yang baik.

### 3.3.6 Interpretation and Evaluation

Pada tahap *interpretation and evaluation*, dimana model dievaluasi dan hasil teknis ditransformasi menjadi *business insights*. Tahap evaluasi menilai kinerja model pada data uji menggunakan empat metrik. *Root Mean Squared Error* digunakan untuk mengukur deviasi prediksi dengan penekanan pada kesalahan besar. *Mean Absolute Error* memberikan ukuran kesalahan absolut rata-rata yang mudah diinterpretasikan. Koefisien determinasi R kuadrat mengukur proporsi variasi lama menginap yang dapat dijelaskan oleh model. *Mean Absolute Percentage Error* menyajikan akurasi dalam persentase relatif. Selain komparasi angka, dilakukan visualisasi hubungan nilai aktual dan prediksi serta distribusi residu guna memverifikasi keberpihakan, sebaran, dan



keberadaan *outlier*. Analisis *feature importance* pada kedua algoritma digunakan untuk mengidentifikasi faktor yang berkontribusi besar terhadap prediksi seperti jumlah tamu, pendapatan hotel, dan pendapatan kamar, sehingga temuan teknis dapat ditautkan dengan konteks operasional. Hasil akhir berupa *business insights* untuk Hotel XYZ.

### 3.3.7 Business Recommendation

Tahap terakhir adalah *business recommendation*, di mana hasil teknis dari model prediktif diubah menjadi *insights* bisnis yang aplikatif serta rekomendasi strategis bagi Hotel XYZ. Tahap akhir menerjemahkan hasil teknis menjadi rekomendasi yang dapat diimplementasikan. Berdasarkan performa terbaik model dan variabel paling berpengaruh, disusun langkah strategis untuk mendorong perencanaan operasional dan promosi musiman hotel. Selain itu, disusun simulasi proyeksi untuk periode mendatang dengan asumsi pertumbuhan realistis pada jumlah tamu, pemesanan, dan pendapatan, sehingga manajemen memiliki dasar perencanaan taktis dan strategis yang terukur.

## 3.4 Teknik Pengumpulan Data

### 3.4.1 Studi Literatur

Studi literatur dalam penelitian ini digunakan untuk membangun fondasi teoretis dan metodologis dalam penerapan algoritma *Random Forest* dan *Gradient Boosting* pada prediksi lama menginap tamu hotel. Melalui kajian literatur, peneliti memperoleh pemahaman tentang parameter optimal dan teknik *preprocessing* yang telah terbukti efektif pada penelitian sebelumnya, sehingga dapat diterapkan untuk meningkatkan akurasi model prediksi pada Hotel XYZ. Selain itu, studi literatur digunakan untuk mengidentifikasi metrik evaluasi yang paling sesuai (RMSE, MAE, MAPE,  $R^2$ ) dalam konteks prediksi regresi hotel. Hasil kajian literatur juga menjadi acuan dalam menentukan *feature engineering* yang tepat dan metodologi *comparative analysis* antara kedua algoritma untuk menghasilkan rekomendasi yang valid bagi manajemen hotel.

### 3.4.2 Populasi Penelitian

Populasi dalam penelitian ini adalah seluruh data transaksi tamu hotel XYZ, dengan periode data mencakup tahun 2023 hingga 2025 yang berjumlah sekitar 357.434 *record*. Populasi ini dipilih karena mencerminkan seluruh spektrum perilaku tamu dan pola reservasi hotel dalam kondisi operasional nyata, termasuk variasi musiman, tipe tamu, dan durasi menginap yang beragam. Karakteristik populasi meliputi data demografi tamu (usia, asal daerah, tipe pelanggan), informasi reservasi (tanggal *check-in*, *check-out*, tipe kamar, harga), serta metadata transaksi lainnya yang terekam dalam sistem *database MariaDB* hotel. Pemilihan populasi ini relevan karena menyediakan dataset yang komprehensif dan representatif untuk mengembangkan model prediksi yang dapat diterapkan secara praktis dalam operasional Hotel XYZ.

### 3.4.3 Sampel Penelitian

Pemilihan sampel dalam penelitian ini dilakukan menggunakan metode *purposive sampling* dengan mempertimbangkan kesesuaian data terhadap tujuan penelitian. Dari total populasi sebanyak 357.434 *record* data reservasi yang tercatat dalam sistem manajemen Hotel XYZ, peneliti melakukan penyaringan berdasarkan periode waktu dan kelengkapan atribut yang relevan dengan pemodelan prediksi. Berdasarkan proses penyaringan dan pembersihan data, diperoleh 2.220 *record* yang digunakan sebagai sampel penelitian. Kriteria pemilihan sampel didasarkan pada aspek temporal (periode data tahun 2024) dan kelengkapan variabel yang dibutuhkan untuk proses analisis. Hal ini dilakukan untuk memastikan data yang dianalisis memiliki karakteristik beragam dan representatif terhadap kondisi operasional hotel selama periode penelitian. Proses ekstraksi sampel dilakukan melalui *Structured Query Language* (SQL) menggunakan *phpMyAdmin*, di mana peneliti secara langsung mengakses *database* Hotel XYZ untuk memperoleh data yang memenuhi kriteria tersebut. Pendekatan ini dipilih agar data yang digunakan bersifat valid, akurat, dan sesuai dengan kebutuhan penelitian, sekaligus meminimalkan potensi bias akibat perbedaan periode waktu atau ketidakkonsistenan data.

### 3.4.4 Metode Pengumpulan Data

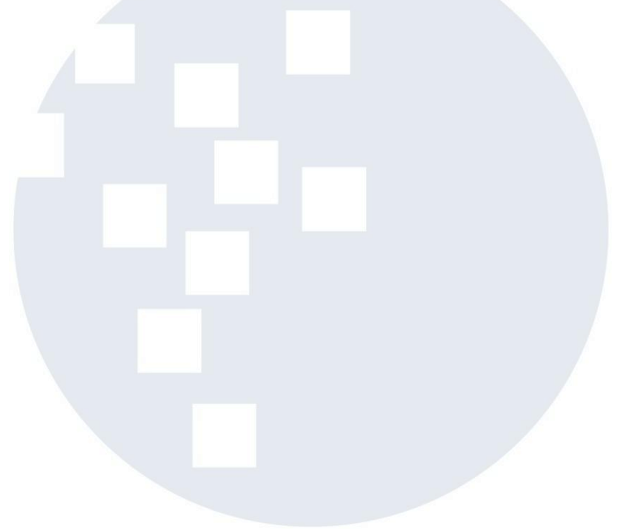
Metode pengumpulan data dalam penelitian ini dilakukan secara langsung melalui akses ke database *internal* milik Hotel XYZ. Data yang digunakan berasal dari sistem manajemen hotel yang terintegrasi dalam database MariaDB, yang merekam seluruh aktivitas reservasi, informasi tamu, serta riwayat durasi menginap secara digital. Pemanfaatan data internal ini dipilih karena Hotel XYZ telah memiliki sistem informasi manajemen yang terstruktur, sehingga memungkinkan peneliti memperoleh data aktual, konsisten, dan relevan dengan kebutuhan analisis. Sebelum proses pengambilan data dilakukan, peneliti berkoordinasi dengan pihak manajemen Hotel XYZ untuk memahami struktur serta konteks penggunaan data, sekaligus mengidentifikasi kemungkinan kendala yang muncul dalam proses akses dan ekstraksi. Melalui diskusi tersebut diperoleh pemahaman bahwa dari keseluruhan database yang besar, hanya sebagian data yang memenuhi kriteria kelengkapan dan konsistensi untuk dijadikan sampel penelitian.

Proses ekstraksi data kemudian dilakukan menggunakan *phpMyAdmin* melalui *query* SQL terstruktur untuk mengambil atribut yang relevan terhadap variabel penelitian, seperti pola reservasi, tanggal *check-in* dan *check-out*, serta jenis kamar yang dipesan. Setiap data yang diambil diverifikasi guna memastikan tidak terdapat duplikasi maupun nilai yang hilang. Setelah itu, data mentah menjalani tahap data *cleaning* dan standarisasi format, termasuk penyesuaian format tanggal, penyebutan tipe kamar, dan status reservasi agar seragam *antarrecord*. Tahapan ini bertujuan untuk meningkatkan kualitas dan keandalan data, sekaligus meminimalkan risiko kesalahan pada proses analisis dan pemodelan prediksi di tahap berikutnya.

### 3.5 Teknik Analisis Data

Teknik analisis data pada penelitian ini dilakukan secara kuantitatif dengan pendekatan statistik regresi dan evaluasi model prediksi. Data yang telah dikumpulkan melalui observasi, dokumentasi, diolah menggunakan tahapan *preprocessing* seperti pembersihan data, *encoding* variabel kategorikal, serta

normalisasi fitur numerik. Selanjutnya, analisis regresi dilakukan dengan menerapkan algoritma *Random Forest* dan *Gradient Boosting* untuk memprediksi lama menginap tamu hotel, di mana variabel *input* meliputi profil tamu, dan detail reservasi. Pengukuran performa model dilakukan menggunakan metrik RMSE, MAE, dan  $R^2$ , yang masing-masing dihitung berdasarkan selisih antara nilai aktual dan prediksi model. Hasil analisis statistik ini kemudian digunakan untuk membandingkan akurasi dan stabilitas kedua algoritma, serta mengidentifikasi variabel yang paling berpengaruh terhadap prediksi lama menginap tamu hotel XYZ.



UMN

UNIVERSITAS  
MULTIMEDIA  
NUSANTARA