BAB III

METODE PENELITIAN

3.1. Pendekatan Penelitian

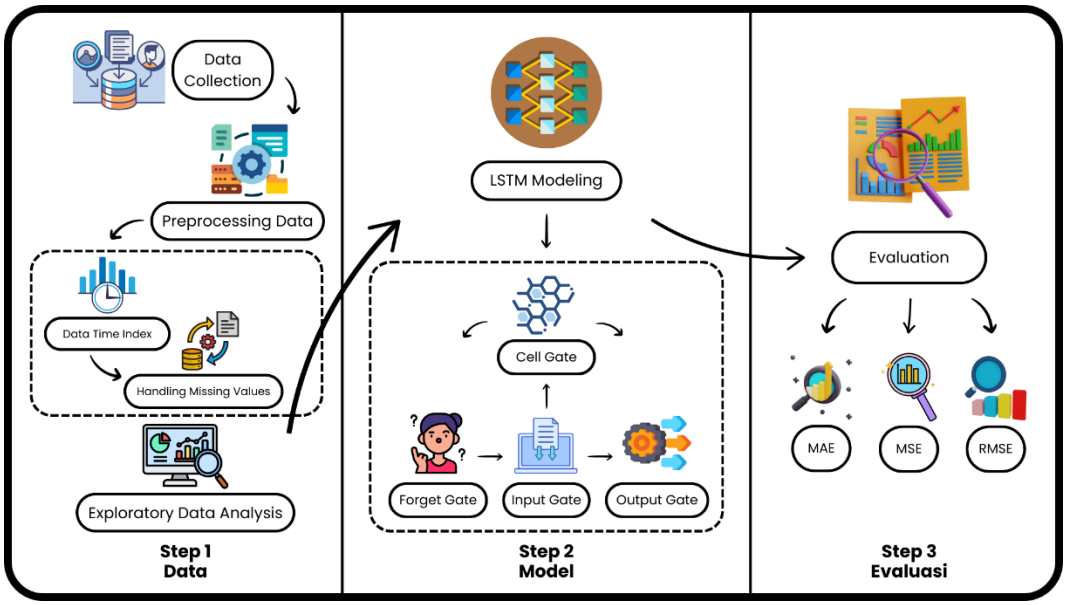
Penelitian ini menggunakan dataset yang diperoleh dari platform Kaggle, dataset ini merepresentasikan data penjualan dari mesin penjual otomatis yang ditempatkan di berbagai lokasi di New Jersey Tengah selama periode Januari 2022 hingga Desember 2022. Dataset ini mencakup informasi transaksi harian, jumlah unit produk yang terjual, serta variabel lain yang mendukung analisis prediksi penyediaan produk.

Proses prediksi dalam penelitian ini dilakukan melalui beberapa tahapan, dimulai dari pengumpulan data, dilanjutkan dengan preprocessing data, termasuk pembersihan dan normalisasi. Selanjutnya, dilakukan pemodelan dengan pendekatan *deep learning* menggunakan arsitektur LSTM, yang dirancang untuk menangani data deret waktu dan mempelajari pola permintaan produk. Model ini dievaluasi menggunakan metrik MAE, MSE, dan RMSE untuk mengukur tingkat akurasi prediksi.

Penelitian ini menggunakan Google Colab berbasis cloud dengan bahasa pemrograman Python. Seluruh proses penelitian, termasuk pengolahan data, pembangunan model, serta evaluasi model, dilakukan menggunakan pustaka seperti TensorFlow, Keras, Pandas, dan Matplotlib. Untuk penyusunan laporan penelitian, digunakan Microsoft Word, serta Mendeley Desktop sebagai alat bantu dalam pengelolaan referensi penelitian.

3.2. Tahapan Penelitian

Tahapan penelitian merupakan kegiatan penelitian yang dilakukan secara terencana, terstuktur, dan sistematis untuk mencapai tujuan tertentu.



Gambar 3. 1 Tahapan Penelitian

Pada Gambar 3.1 merupakan tahapan penelitan yang akan dilakukan melalui tiga tahapan utama. Diantaranya tahapan pertama adalah pengumpulan dan pra-pemrosesan data, di mana data yang relevan dikumpulkan, dibersihkan, dan disiapkan untuk analisis lebih lanjut. Tahap kedua melibatkan pembuatan model yang di mana algoritma atau arsitektur yang sesuai dirancang dan diimplementasikan untuk memproses data tersebut. Selanjutnya tahapan terakhir adalah evaluasi model, di mana kinerja model diukur menggunakan metrik pengukuran yang menggambarkan persentase dari kesalahan peramalan. Ketiga tahapan ini dilakukan secara sistematis untuk mencapai tujuan penelitian dengan hasil yang diharapkan.

3.2.1. *Data Collection*

Dataset yang digunakan dalam penelitian ini diperoleh dari platform *Kaggle* dengan tautan https://www.kaggle.com/datasets/awesomeasingh/vending-machine-sales/data. Dataset ini merepresentasikan data penjualan dari mesin penjual otomatis yang ditempatkan di berbagai lokasi di New Jersey Tengah selama periode Januari 2022 hingga Desember 2022.

Dataset ini mencakup informasi seperti lokasi mesin penjual otomatis, volume penjualan, serta data waktu dalam skala harian selama satu tahun penuh. Data ini akan digunakan sebagai bahan utama untuk membangun model prediksi penyediaan produk dengan memanfaatkan pendekatan *deep learning* dengan algoritma LSTM.

3.2.2. *Data Preprocessing*

Pada tahap ini, data yang telah dikumpulkan akan diproses agar siap digunakan dalam analisis dan pemodelan. Langkah pertama yang dilakukan adalah konversi kolom *transdate* ke dalam format *datetime*, sehingga data dapat dikenali sebagai deret waktu (*time series*). Setelah berhasil dikonversi, kolom *transdate* kemudian ditetapkan sebagai indeks utama dalam struktur *dataframe*. Langkah ini penting untuk mendukung pengurutan data secara kronologis dan memungkinkan model prediktif untuk mempelajari pola historis berdasarkan urutan waktu yang tepat.

Selanjutnya dilakukan ekstraksi fitur temporal tambahan dari kolom *transdate*, yaitu atribut *month* dan *day*. Penambahan fitur ini bertujuan untuk memperkaya informasi waktu dan mendukung analisis pola musiman atau fluktuasi permintaan berdasarkan hari atau bulan tertentu. Langkah ini juga membantu model dalam mengenali tren musiman yang mungkin muncul dalam data penjualan. Setelah itu, dilakukan proses normalisasi teks pada kolom *Category* dengan mengubah seluruh

nilai menjadi huruf kecil guna menjaga konsistensi penamaan kategori produk, terutama saat dilakukan proses pengelompokan atau pemfilteran dengan menambahkan dua fitur biner baru yaitu *food* dan *drink*. Penambahan ini digunakan untuk menandai apakah suatu produk tergolong ke dalam kategori makanan atau minuman. Transformasi ini berguna dalam klasifikasi produk serta membantu model memahami jenis barang yang dijual.

Tahapan selanjutnya adalah penanganan data hilang *(handling missing values).* Proses ini diawali dengan mengidentifikasi baris yang memiliki nilai kosong pada fitur penting, khususnya kolom *Category* yang berperan dalam klasifikasi dan segmentasi produk. Baris – baris dengan nilai kosong pada kolom tersebut dihapus untuk menjaga kualitas dan integritas data. Hasil dari tahap ini adalah dataset yang lebih bersih, konsisten, dan siap untuk digunakan dalam proses pelatihan model prediksi permintaan produk.

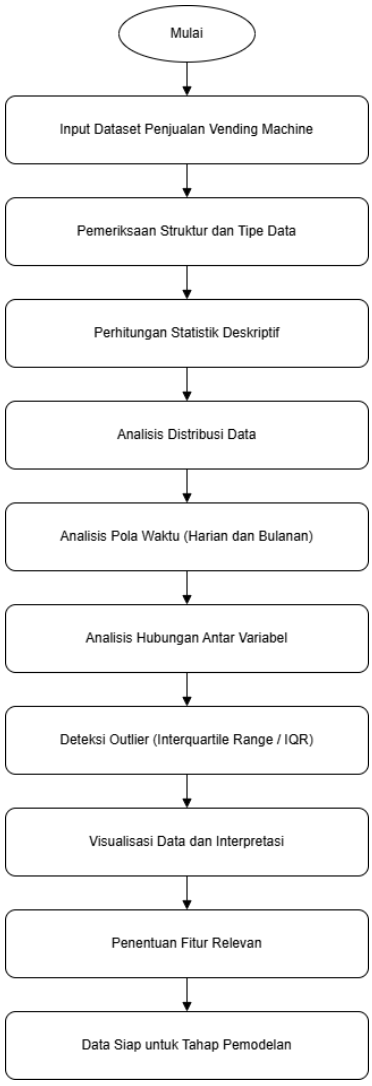
3.2.3. *Exploratory Data Analysis* (EDA)

Exploratory Data Analysis (EDA) merupakan tahap awal yang dilakukan untuk memperoleh pemahaman awal terhadap struktur dan karakteristik data penjualan pada *vending machine*. Tahapan ini bertujuan untuk mengidentifikasi pola distribusi, tren penjualan, pola musiman, hubungan antar variabel, serta keberadaan data ekstrem (*outlier*) yang berpotensi memengaruhi kinerja model prediksi. Analisis dilakukan terhadap data penjualan berdasarkan skala waktu bulanan dan harian, guna melihat pengaruh periode waktu terhadap volume penjualan masing-masing kategori produk yaitu makanan dan minuman.

Proses EDA dilengkapi dengan visualisasi grafik untuk membantu menginterpretasikan tren yang muncul, baik dalam skala bulanan maupun harian.

Hasil dari analisis visual ini memberikan wawasan awal terhadap perilaku konsumen dan digunakan sebagai dasar dalam perancangan model prediksi permintaan yang lebih akurat dan responsif.

Penerapan EDA dalam penelitian ini mengacu pada arsitektur EDA yang telah dijelaskan pada BAB II. Arsitektur tersebut disusun sebagai kerangka analisis data yang terstruktur dan sistematis untuk memastikan setiap tahapan eksplorasi dilakukan secara konsisten dan mendukung tujuan penelitian. Untuk memperjelas alur penerapan EDA dalam metodologi penelitian, tahapan analisis divisualisasikan dalam bentuk *flowchart* sebagaimana ditunjukkan pada Gambar 3.2.



Gambar 3. 2 *Flowchart Exploratory Data Analysis* (EDA)

Berdasarkan Gambar 3.2 *Exploratory Data Analysis* (EDA) dilakukan secara bertahap untuk memastikan data penjualan siap digunakan pada proses pemodelan. Tahapan EDA diawali dengan pemeriksaan struktur dan tipe data untuk menjamin kesesuaian format dan kelengkapan data, kemudian dilanjutkan dengan perhitungan statistik deskriptif dan analisis distribusi guna memahami karakteristik serta sebaran nilai penjualan. Selanjutnya analisis pola waktu dilakukan pada skala harian dan bulanan untuk mengidentifikasi tren dan pola musiman, diikuti dengan analisis hubungan antarvariabel untuk mengetahui keterkaitan antara waktu dan volume penjualan. Proses EDA juga mencakup deteksi *outlier* untuk mengidentifikasi data ekstrem yang berpotensi memengaruhi kinerja model, serta visualisasi data untuk memperkuat interpretasi pola dan tren yang muncul. Tahap akhir EDA adalah penentuan fitur relevan yang akan digunakan pada proses pemodelan LSTM, sehingga data yang dihasilkan telah terstruktur dan siap digunakan pada tahap selanjutnya.

3.2.4. *LSTM Modeling*

Penerapan Long Short-Term Memory (LSTM) dalam penelitian ini bertujuan untuk memprediksi permintaan produk berdasarkan pola historis penjualan *vending machine*. LSTM dipilih karena kemampuannya dalam menangani data deret waktu serta mempertahankan informasi jangka panjang yang diperlukan untuk menangkap tren musiman dan pola fluktuasi permintaan. Model ini bekerja dengan menggunakan sel memori yang dapat mengontrol informasi mana yang harus dipertahankan atau dilupakan pada setiap langkah waktu.

Pemodelan diawali dengan pembagian data (*data splitting*) menjadi data training dan data testing untuk memastikan model dapat belajar pola dari data

historis serta diuji performanya pada data yang belum pernah dilihat sebelumnya. Selanjutnya, data historis yang telah melalui preprocessing dikonversi menjadi format sekuensial agar dapat digunakan sebagai input dalam model LSTM. Setiap data masukan terdiri dari sejumlah hari sebelumnya sebagai fitur, sedangkan target keluaran adalah prediksi jumlah penyediaan produk untuk periode berikutnya. Proses kerja LSTM dalam memproses input mengikuti mekanisme berikut:

1. Forget Gate

   Pada tahap ini, model memutuskan informasi apa yang perlu dibuang dari sel memori berdasarkan input saat ini (x_t) dan status tersembunyi sebelumnya (s_{t-1}). Fungsi aktivasi sigmoid digunakan untuk menghasilkan nilai antara 0 dan 1, yang menunjukkan seberapa besar bagian dari memori sebelumnya yang akan dilupakan.

2. Input Gate

   Input gate dilakukan dengan mengalikan output dari input gate dengan kandidat memori baru, yang dihitung menggunakan fungsi aktivasi tanh.

3. Cell Gate

   Nilai sel memori diperbarui dengan menggabungkan informasi lama yang dipertahankan oleh forget gate dan informasi baru dari input gate.

4. Output Gate

   Setelah sel memori diperbarui, model menentukan output akhir berdasarkan memori yang telah diperbarui. Fungsi sigmoid digunakan untuk menentukan seberapa banyak informasi yang akan diambil dari sel memori untuk menghasilkan output

Output (h_{t-1}) ini kemudian digunakan sebagai prediksi penyediaan produk untuk periode berikutnya. Untuk memastikan bahwa model menghasilkan prediksi yang akurat, dilakukan training dengan memberikan sejumlah data input (penjualan beberapa hari sebelumnya) dan target output (penjualan pada hari berikutnya).

3.2.5. Evaluation

Evaluasi model dilakukan untuk menilai seberapa baik model LSTM dalam memprediksi penyediaan produk pada *vending machine*. Setelah model LSTM selesai dilatih menggunakan data training, tahap selanjutnya adalah testing. Testing dilakukan untuk mengukur kemampuan model LSTM dalam memprediksi penyediaan produk pada data yang belum pernah dilihat sebelumnya. Dalam proses testing, model menerima input dari x_test, yaitu sekumpulan data penjualan *vending machine* pada periode tertentu. Model kemudian menghasilkan output berupa prediksi nilai y_pred, yang merepresentasikan jumlah stok produk yang diprediksi untuk periode berikutnya berdasarkan pola historis. Hasil prediksi ini kemudian dibandingkan dengan nilai aktual y_test, yang merupakan jumlah stok produk yang benar-benar terjadi pada periode tersebut. Untuk mengukur performa model pada tahap evaluasi, digunakan berbagai metrik evaluasi, seperti:

1. *Mean Absolute Error* (MAE), mengukur rata-rata selisih absolut antara nilai aktual dan prediksi.
2. *Mean Squared Error* (MSE), menghitung rata-rata selisih kuadrat antara nilai aktual dan prediksi.
3. *Root Mean Squared Error* (RMSE), memberikan nilai error dalam skala yang sama dengan data asli.

Selain itu, proses evaluasi juga mencakup analisis visualisasi hasil prediksi terhadap data aktual dalam bentuk grafik, untuk melihat apakah model mampu mengikuti pola tren data historis. Jika nilai error pada testing tinggi, model dapat mengalami overfitting atau underfitting, sehingga perlu dilakukan penyesuaian seperti tuning hyperparameter atau penambahan fitur. Melalui tahap evaluasi ini, diharapkan model mampu menggeneralisasi pola dari data historis dan memberikan prediksi penyediaan yang akurat untuk mendukung pengelolaan *vending machine* secara lebih efisien dan mengurangi risiko kehabisan atau kelebihan stok.