

BAB III

METODOLOGI PENELITIAN

3.1 Objek Penelitian

Objek penelitian yang penelitian kali ini adalah *real case* yang terdapat pada perusahaan asuransi. Pada penelitian ini data yang diberikan dari perusahaan merupakan jumlah penjualan produk asuransi perusahaan perharinya yang didapatkan dari beberapa channel yang dimiliki. Data yang dibagikan ini merupakan data historis dari tanggal 1 Januari 2024 Hingga 19 Maret 2025 yang berisi informasi mengenai total jumlah penjualan dan tanggal dari penjualan tersebut.

Tabel 3.1 Isi Dataset

No.	Column	Tipe Data	Deskripsi
1	TRANS_DATE	DateTime[ns]	Tanggal masuknya data transaksi kedalam database perusahaan.
2.	DAILY_GWP_CASE	Int64	Jumlah data penjualan dari perusahaan yang didapatkan dalam periode waktu tertentu dan dari channel tertentu

Dataset yang didapatkan ini sebelumnya sudah diolah dan diberikan untuk penelitian sehingga beberapa informasi seperti channel yang memberikan informasi ini tidak ditunjukkan. Perlu di informasikan bahwa data yang diberikan sudah dikhususkan dari perusahaan untuk memenuhi kebutuhan penelitian ini.

3.2 Metode Penelitian

Dalam penelitian ini, pendekatan yang digunakan adalah metode CRISP-DM. Proses penelitian dimulai dengan tahap pemahaman masalah (*business understanding*), yang bertujuan untuk mengidentifikasi kebutuhan divisi manajemen dalam menganalisis penjualan produk. Selanjutnya, tahap pemahaman data (*data understanding*) dilakukan dengan menganalisis data historis penjualan

guna memahami pola serta tren yang terjadi. Setelah itu, pada tahap persiapan data (*data preparation*), data melalui proses pembersihan, transformasi, serta pembagian ke dalam set pelatihan dan pengujian agar siap untuk tahap pemodelan.

Pada tahap pemodelan (*modeling*), algoritma ARIMA dan LSTM digunakan untuk membangun model prediksi penjualan berbasis *time series*. Model yang telah dikembangkan kemudian dievaluasi pada tahap evaluasi (*evaluation*) dengan mengukur tingkat kesalahan menggunakan metrik MAE, MSE, dan R^2 menilai akurasi prediksi. Untuk mendapatkan model yang terbaik, maka dilakukan tahap validasi yang akan menguji pemodelan berdasarkan pembagian data. Pemodelan yang terbaik kemudian didapatkan dan digunakan untuk tahap deployment.

Akhirnya, pada tahap deployment, hasil penelitian diimplementasikan dalam bentuk *dashboard*, yang menampilkan data perusahaan serta hasil prediksi yang telah dihasilkan oleh model.

Berikut merupakan rincian mengenai metode penelitian yang digunakan.

Tabel 3.2 Tahapan Metode Penelitian

No.	Tahap Metode	Penjelasan
1	Business Understanding	Memahami kebutuhan analisa dari divisi <i>Sales</i> perusahaan dan juga hasil prediksi yang diinginkan untuk kemajuan dan produksi perusahaan
2	Data Understanding	Menganalisa pola dari data historis penjualan asuransi selama 2 tahun kebelakang
3	Data Preparation	Membersikan dan mengubah format data yang disesuaikan untuk dapat dilakukannya serta pembagian data menjadi data pelatihan dan data pengujian untuk pemodelan <i>machine learning</i> di tahap berikutnya

4	Data Modeling	Melakukan pemodelan algoritma ARIMA dan LSTM berdasarkan data yang sudah diolah
5	Data Evaluation	Mengevaluasi hasil pemodelan <i>machine learning</i> yang dilakukan menggunakan algoritma MAE, MSE, dan R^2 untuk melihat keakuratan hasil dari prediksi. Melakukan Validasi model dengan variasi ratio pembagian data train-test.
6	Deployment	Pembuatan <i>dashboard</i> sederhana untuk menampilkan data dan juga hasil prediksi

3.3 Variable Penelitian

Dalam penelitian ini, variable penelitian yang digunakan terdiri dari 2 jenis yaitu variable dependen dan variable independen. Variable dependen merupakan variable penyebab yang mempengaruhi atau berpengaruh dalam melakukan prediksi penjualan. Sementara variable independen merupakan variable akibat atau hasil angka penjualan dalam waktu yang terpengaruh dari variable dependen. Dalam melakukan prediksi, penelitian ini hanya mendapatkan 2 kolom data yang dilakukan pemodelan berdasarkan data historis.

3.3.1 Variable Independen

Dalam melakukan prediksi, variable independen penelitian ini hanya menggunakan 1 variable atau kolom yaitu data tanggal “TRANS_DATE”. Variable ini dibutuhkan untuk menunjukkan waktu masuknya data penjualan kedalam database sehingga menjadi sebuah data historis.

3.3.2 Variable Dependen

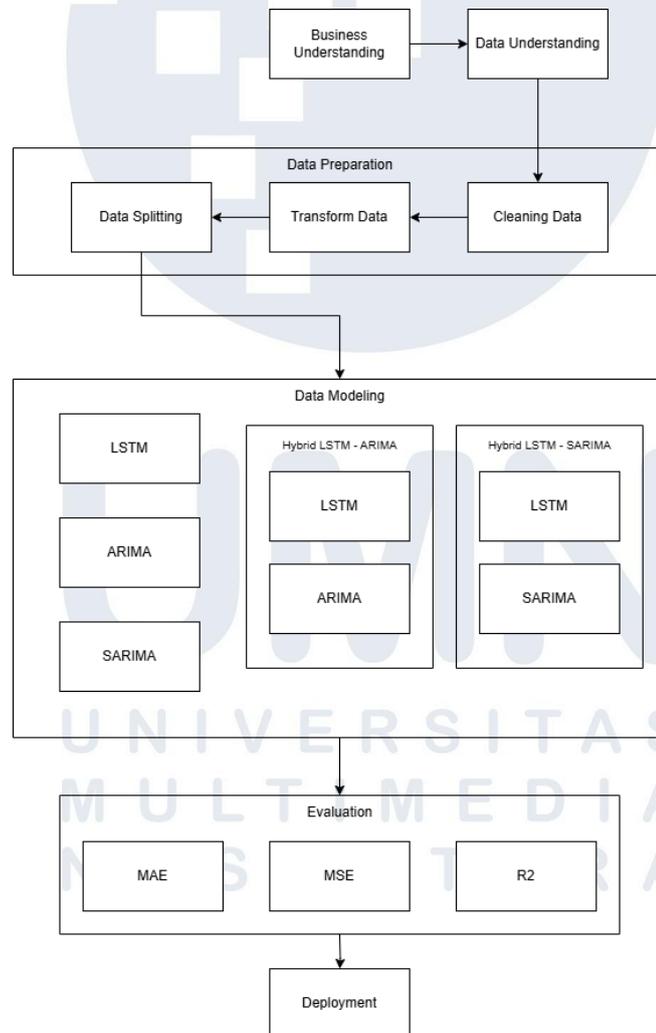
Data yang digunakan sebagai variable dependen penelitian ini adalah variabel atau data kolom “DAILY_GWP_CASE” yaitu data penjualan per tanggal. Isi data ini adalah jumlah terjualnya produk perusahaan dalam satu waktu/tanggal tersebut.

3.4 Teknik Pengumpulan Data

Penelitian ini menggunakan data yang diberikan dari perusahaan. Data dari perusahaan berisi penjualan per harinya milik perusahaan. Data ini terdiri dari 2 kolom yaitu kolom angka *Sales*, dan jumlah penjualan atau *case* yang didapatkan dalam waktu tersebut.

3.5 Teknik Analisis Data

Penelitian ini menerapkan pendekatan CRISP-DM, yang merupakan standar umum dalam proses data mining. Metode ini terdiri dari enam tahap utama, yaitu *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation*, dan *Deployment*.



Gambar 3.1 Diagram Teknik Analisis Data

Berikut adalah penjelasan rinci mengenai teknik analisis data yang diterapkan dalam penelitian ini:

3.5.1 Business Understanding

Tahap ini bertujuan untuk memahami kebutuhan bisnis dan menentukan tujuan utama dari penelitian. Dalam konteks penelitian ini, pemahaman masalah berfokus pada kebutuhan divisi manajemen dalam menganalisis pola penjualan serta mendapatkan prediksi penjualan yang akurat. Hasil dari tahap ini adalah pemetaan kebutuhan analisis yang relevan dengan kepentingan bisnis.

3.5.2 Data Understanding

Pada tahap ini, dilakukan eksplorasi terhadap data historis penjualan guna memahami pola serta tren yang terjadi dalam rentang waktu tertentu. Data yang digunakan dalam penelitian ini adalah data penjualan produk asuransi selama dua tahun terakhir. Analisis awal meliputi pemeriksaan distribusi data, tren, serta deteksi anomali yang mungkin terjadi dalam data.

3.5.3 Data Preparation

Tahap ini mencakup serangkaian proses yang bertujuan untuk mempersiapkan data agar dapat digunakan dalam tahap pemodelan. Proses yang dilakukan meliputi:

- 1) **Pembersihan Data:** Menghapus data yang tidak relevan, menangani nilai yang hilang (*missing values*), serta mengatasi duplikasi data.
- 2) **Transformasi Data:** Mengubah format data agar sesuai dengan kebutuhan algoritma prediksi, termasuk normalisasi data jika diperlukan.
- 3) **Pembagian Data:** Data dibagi menjadi *training set* dan *testing set* dengan rasio 80:20 untuk memastikan model dapat belajar dengan baik dan diuji secara objektif.

3.5.4 Data Modeling

Pada tahap ini, dilakukan pembangunan model prediksi berbasis *time series* menggunakan dua algoritma utama, yaitu:

- 1) *ARIMA (Auto Regressive Integrated Moving Average)*: Algoritma statistik yang digunakan untuk memodelkan data deret waktu dengan mempertimbangkan aspek *autoregressive*, *moving average*, dan diferensiasi data guna menangkap tren serta pola musiman.
- 2) *SARIMA (Seasonal ARIMA)*: Pengembangan dari ARIMA yang secara eksplisit memperhitungkan komponen musiman dalam data deret waktu. SARIMA menambahkan parameter musiman pada model ARIMA, sehingga lebih efektif dalam memodelkan data yang memiliki pola berulang secara periodik, seperti mingguan, bulanan, atau tahunan.
- 3) *LSTM (Long Short-Term Memory)*: Jenis *recurrent neural network (RNN)* yang mampu menangkap hubungan jangka panjang dalam data *time series* dan secara efektif mengatasi permasalahan *vanishing gradient*, sehingga lebih unggul dalam memprediksi pola kompleks.

3.5.5 Evaluation

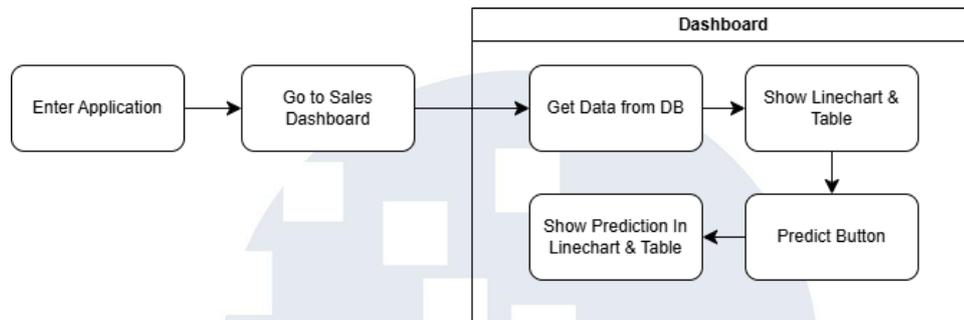
Model yang telah dibangun dievaluasi dengan menggunakan tiga metrik utama:

- 1) *Mean Absolute Error (MAE)*: Mengukur rata-rata kesalahan absolut antara nilai prediksi dan nilai aktual.
- 2) *Mean Squared Error (MSE)*: Menghitung rata-rata kesalahan kuadrat, memberikan bobot lebih besar pada kesalahan yang lebih besar.
- 3) *R-squared (R²)*: Mengukur proporsi variansi dalam data aktual yang dapat dijelaskan oleh model. Nilai R² berada dalam rentang 0 hingga 1, di mana nilai yang lebih tinggi menunjukkan bahwa model dapat menjelaskan variabilitas data dengan lebih baik.

Tahap evaluasi ini bertujuan untuk memastikan bahwa model yang digunakan memiliki performa yang optimal dan dapat diandalkan dalam memprediksi penjualan di masa mendatang.

3.5.6 Deployment

Tahap akhir dari penelitian ini adalah implementasi hasil prediksi ke dalam bentuk *dashboard* yang dapat diakses oleh divisi terkait di perusahaan. *Dashboard* ini berfungsi sebagai alat bantu visualisasi yang menampilkan data historis, pola penjualan, serta prediksi hasil dari model LSTM.



Gambar 3.2 Diagram Deployment *Dashboard*

Sistem *dashboard* prediksi penjualan dikembangkan dengan alur yang terstruktur seperti yang ditunjukkan pada diagram. Pengguna terlebih dahulu masuk ke dalam aplikasi dan mengakses halaman *Sales Dashboard*. Selanjutnya, sistem akan mengambil data historis penjualan dari database dan menampilkannya dalam bentuk grafik garis dan tabel. Pengguna kemudian dapat menekan tombol *Predict* untuk menjalankan model prediksi penjualan berbasis *machine learning time series*. Hasil prediksi akan langsung ditampilkan secara visual dalam grafik dan tabel yang sama, sehingga pengguna dapat membandingkan data aktual dan prediksi secara *real-time* untuk mendukung pengambilan keputusan yang cepat dan tepat.

Dengan mengikuti tahapan CRISP-DM ini, penelitian ini diharapkan dapat menghasilkan sistem prediksi penjualan yang akurat serta mendukung pengambilan keputusan bisnis yang lebih baik.