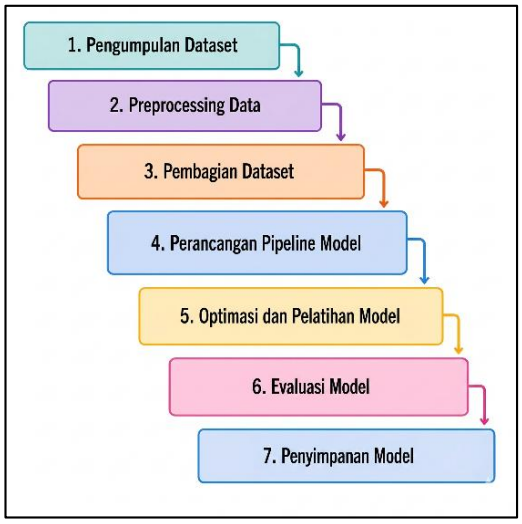
BAB III
METODOLOGI PENELITIAN

3.1. Prosedur Penelitian

Prosedur penelitian disusun secara sistematis sebagai pedoman operasional untuk mencapai tujuan penelitian. Penelitian diawali dengan pengumpulan dataset yang memuat parameter *Pressure* dan *Flow_Rate*, kemudian dilanjutkan dengan proses *preprocessing* untuk meningkatkan kualitas data. Dataset yang telah diproses selanjutnya dibagi menjadi data *training* dan *testing*, kemudian digunakan dalam pembentukan *pipeline* model Logistic Regression dengan *Polynomial Features*. Tahap berikutnya evaluasi performa model, serta penyimpanan model terbaik ke dalam format *.pkl*. Tahapan prosedur penelitian yang dilakukan dapat dilihat pada Gambar 3.1.



Gambar 3.1. Tahapan Prosedur Penelitian

3.1.1. Pengumpulan Dataset

Tahap pengumpulan *dataset* dilakukan dengan menggabungkan satu *dataset* utama dan dua *dataset* hasil augmentasi. *Dataset* yang digunakan memuat parameter *Pressure* dan *Flow_Rate* sebagai variabel masukan, serta *Leakage_Flag* sebagai label keluaran yang menunjukkan kondisi pipa normal atau bocor. Penggabungan *dataset* utama dengan *dataset* augmentasi dilakukan untuk menambah jumlah data dan memperluas variasi pola data yang digunakan dalam proses *training* model *machine learning*. Data yang telah digabungkan kemudian digunakan sebagai dasar dalam proses klasifikasi kondisi pipa menggunakan metode *Logistic Regression* dengan *Polynomial Features*.

3.1.2. Prepocessing Data

Preprocessing data dilakukan untuk memastikan *dataset* yang digunakan sudah bersih dan layak diproses pada tahap *training* model. Tahap ini penting karena kualitas data sangat memengaruhi hasil pembelajaran model *Machine learning*. Data yang masih mengandung nilai kosong, duplikasi, tipe data yang tidak sesuai, atau *Outlier* dapat mengganggu proses *training* dan menurunkan performa model. Pada penelitian ini, *Preprocessing* data dilakukan melalui beberapa tahapan sebagai berikut:

1. Penghapusan Data Kosong atau *Missing value.*

   Penghapusan data kosong dilakukan untuk menghilangkan data yang tidak memiliki nilai pada atribut tertentu. Data kosong perlu dihapus karena dapat menyebabkan kesalahan pada proses pengolahan data dan *training* model.

2. Penghapusan Data Duplikat.

   Penghapusan data duplikat dilakukan untuk menghilangkan data yang memiliki nilai sama secara berulang. Tahap ini bertujuan agar data yang sama tidak memberikan pengaruh berlebihan terhadap proses pembelajaran model.

3. Konversi Data ke Format Numerik

   Konversi data dilakukan pada atribut *Pressure*, *Flow_Rate*, dan *Leakage_Flag* agar seluruh data berada dalam bentuk numerik. Hal ini diperlukan karena algoritma *Machine learning* hanya dapat memproses data dalam bentuk angka.

4. Penghapusan *Outlier* Menggunakan Metode *Z-Score*

   Penghapusan *Outlier* dilakukan untuk menghilangkan data yang memiliki nilai terlalu jauh dari pola umum *dataset*. Pada penelitian ini, metode *Z-Score* digunakan dengan batas nilai kurang dari 3. Data yang memiliki nilai *Z-Score* melebihi batas tersebut dianggap sebagai *Outlier* dan dihapus dari *dataset*.

5. *Smoothing* Data Menggunakan *Moving Average*

   Smoothing data dilakukan untuk mengurangi fluktuasi nilai yang terlalu tajam pada atribut *Pressure* dan *Flow_Rate*. Pada penelitian ini, *Smoothing* dilakukan menggunakan metode *Moving Average* dengan ukuran *window* sebesar 3, sehingga nilai data menjadi lebih stabil.

6. *Cleaning* Ulang Setelah Proses Smoothing

   Cleaning ulang dilakukan setelah proses *Smoothing* untuk memastikan *dataset* tetap bersih dan konsisten. Tahap ini bertujuan memastikan tidak terdapat data kosong, data tidak valid, atau perubahan struktur data setelah proses *Moving Average* diterapkan.

3.1.3. Pembagian Dataset

Dataset yang telah melalui tahap *Preprocessing* kemudian dibagi menjadi data *training* dan data *testing*. Data *training* digunakan untuk melatih model agar mampu mempelajari pola hubungan antara tekanan air dan debit air terhadap kondisi pipa. Sementara itu, data *testing* digunakan untuk menguji kemampuan model dalam mengklasifikasikan data baru yang tidak digunakan pada proses *training*.

Pada penelitian ini, pembagian dataset dilakukan menggunakan metode *train_test_split* dengan proporsi 80% untuk data *training* dan 20% untuk data *testing*. Dengan pembagian tersebut, model dapat

dilatih menggunakan sebagian besar data dan diuji menggunakan data yang berbeda, sehingga kemampuan generalisasi model dapat diketahui secara lebih objektif.

3.1.4. Pembentukan Pipeline Model

Pembentukan model dilakukan menggunakan *pipeline* yang terdiri dari beberapa tahapan, yaitu *Polynomial Features*, *StandardScaler*, *SMOTE*, dan *Logistic Regression*. *Pipeline* digunakan agar proses transformasi fitur, standardisasi data, penyeimbangan kelas, dan *training* model dapat berjalan dalam satu alur yang terstruktur. Dengan penggunaan *pipeline*, setiap tahapan pemodelan dapat dilakukan secara konsisten mulai dari data masukan hingga menghasilkan prediksi kondisi pipa. Tahapan dalam *pipeline* model pada penelitian ini adalah sebagai berikut.

1. *Polynomial Features*

   Polynomial Features digunakan untuk membentuk fitur baru dari atribut *Pressure* dan *Flow_Rate*. Pembentukan fitur *polinomial* bertujuan agar model dapat mengenali hubungan *non-linear* antara tekanan air dan debit air. Dengan adanya fitur *polinomial*, model tidak hanya mempelajari hubungan sederhana antarvariabel, tetapi juga dapat mempelajari kombinasi fitur yang lebih kompleks.

2. *StandardScaler*

   StandardScaler digunakan untuk menyamakan skala nilai pada setiap fitur. Proses standardisasi diperlukan karena fitur hasil transformasi *polinomial* dapat memiliki rentang nilai yang berbeda-beda. Dengan standardisasi, setiap fitur memiliki skala yang lebih seimbang sehingga proses *training* model dapat berjalan lebih optimal.

3. *SMOTE*

   SMOTE digunakan untuk menyeimbangkan distribusi kelas pada data *training*. Teknik ini bekerja dengan membentuk data sintetis pada kelas minoritas agar jumlah data antar kelas menjadi lebih seimbang. Dengan adanya *SMOTE*, model tidak hanya cenderung mempelajari kelas yang jumlah datanya lebih dominan, tetapi juga dapat mengenali pola pada kelas minoritas.

4. *Logistic Regression*

   Logistic Regression digunakan sebagai algoritma utama untuk melakukan klasifikasi kondisi pipa berdasarkan label *Leakage_Flag*. Algoritma ini mempelajari hubungan antara fitur masukan, yaitu *Pressure* dan *Flow_Rate*, dengan target klasifikasi berupa kondisi pipa Normal atau Bocor.

3.1.5. Optimasi dan Training Model

Optimasi dan *training* model merupakan tahap untuk membangun model klasifikasi kondisi pipa berdasarkan data *training*.

Model yang telah dibentuk pada tahap sebelumnya kemudian dilatih menggunakan metode *GridSearchCV* untuk memperoleh kombinasi *hyperparameter* terbaik.

Proses optimasi dilakukan dengan menerapkan *5-fold Cross Validation* serta menguji beberapa parameter, yaitu derajat *polynomial* (degree), nilai regularisasi (C), jenis *solver*, dan jumlah maksimum iterasi (max_iter). Setelah proses optimasi selesai dilakukan, model dengan kombinasi parameter terbaik digunakan sebagai model akhir untuk tahap evaluasi menggunakan data testing.

3.1.6. Evaluasi Model

Evaluasi model dilakukan untuk mengetahui kemampuan model *Logistic Regression* dengan *Polynomial Features* dalam mengklasifikasikan kondisi pipa. Pada tahap ini, model diuji menggunakan data *testing* yang sebelumnya tidak digunakan dalam proses *training*. Pengujian dengan data *testing* bertujuan untuk mengetahui seberapa baik model dalam memprediksi data baru.

Metrik evaluasi yang digunakan dalam penelitian ini meliputi *Accuracy, Precision, Recall, F1-Score, confusion matrix*. *Accuracy* digunakan untuk mengetahui tingkat ketepatan prediksi secara keseluruhan. *Precision* digunakan untuk mengukur ketepatan model dalam memprediksi suatu kelas. *Recall* digunakan untuk mengetahui kemampuan model dalam menemukan data pada kelas tertentu. *F1-Score* digunakan untuk melihat keseimbangan antara *Precision* dan

Recall. Selain itu, *confusion matrix* digunakan untuk menampilkan jumlah prediksi benar dan salah pada masing-masing kelas. Evaluasi ini dilakukan agar performa model dapat dianalisis secara menyeluruh, baik dari segi ketepatan prediksi maupun kemampuan model dalam membedakan kedua kelas.

3.1.7. Penyimpanan Model

Setelah proses *training*, optimasi, dan evaluasi selesai dilakukan, model terbaik disimpan dalam format *.pkl*. File model yang dihasilkan pada penelitian ini adalah *logisticregresion.pkl*. Penyimpanan model dilakukan agar hasil *training* dapat digunakan kembali pada tahap pengujian berikutnya serta memudahkan proses penerapan model tanpa harus mengulang seluruh tahapan *training* dari awal.

3.2. Metode Pengumpulan Data

Metode pengumpulan data dilakukan untuk memperoleh data dan informasi yang dibutuhkan dalam penelitian. Metode pengumpulan data yang digunakan meliputi observasi, wawancara, dan studi literatur.

3.2.1 Observasi

Observasi dilakukan dengan mengunjungi lokasi penelitian, yaitu Perumda Air Minum Tirta Bahari Kota Tegal, untuk memperoleh informasi mengenai kondisi jaringan pipa distribusi dan sistem pemantauan air yang sedang berjalan. Observasi ini bertujuan

untuk memahami proses pemantauan debit dan tekanan air, mengidentifikasi potensi gangguan pada jaringan pipa seperti penurunan tekanan atau perubahan debit air yang tidak normal, serta mengetahui kebutuhan data yang relevan dengan proses pemodelan *machine learning*.

Data utama yang digunakan dalam penelitian ini diperoleh dari *dataset* sekunder *Kaggle*, yaitu *Smart Water Leak Detection Dataset*. *Dataset* tersebut digunakan karena memiliki parameter yang sesuai dengan kebutuhan penelitian, yaitu *Pressure* sebagai tekanan air, *Flow_Rate* sebagai debit air, dan *Leakage_Flag* sebagai label kondisi pipa. Hasil observasi lapangan digunakan sebagai dasar pendukung untuk memahami kondisi permasalahan di lapangan, sedangkan *dataset kaggle* digunakan sebagai data *training* dan pengujian model *Logistic Regression* dengan *Polynomial Features* dalam mengklasifikasikan kondisi pipa menjadi Normal dan Bocor.

3.2.2 Wawancara

Wawancara dilaksanakan kepada pihak teknis atau bagian distribusi Perumda Air Minum Tirta Bahari Kota Tegal sebagai narasumber penelitian. Teknik wawancara yang digunakan adalah semi-terstruktur, yaitu peneliti menyiapkan daftar pertanyaan pokok mengenai pemantauan debit dan tekanan air, tetapi tetap memberikan ruang kepada narasumber untuk menyampaikan penjelasan tambahan berdasarkan kondisi nyata di lapangan.

Pelaksanaan wawancara ini bertujuan untuk memperoleh informasi pendukung mengenai proses pemantauan jaringan pipa distribusi, kendala yang terjadi dalam pengukuran debit dan tekanan air, serta permasalahan yang berpotensi mengarah pada kondisi kebocoran pipa. Hasil wawancara kemudian digunakan sebagai acuan dalam memahami kebutuhan penelitian dan mendukung penerapan model *machine learning* berbasis *Logistic Regression* dengan *Polynomial Features* untuk klasifikasi kondisi pipa Normal dan Bocor. Kegiatan observasi dan wawancara ditunjukkan pada Gambar 3.2.

Gambar 3.2. Observasi Wawancara

3.2.3 Studi Literatur

Studi literatur dilakukan dengan mengumpulkan berbagai referensi berupa jurnal ilmiah, artikel *website* lainnya yang berkaitan dengan deteksi kebocoran pipa, tekanan air, debit air, *machine learning*, serta metode *Logistic Regression* dengan *Polynomial Features*. Selain itu, studi literatur juga digunakan untuk memahami proses *Preprocessing* data, pembentukan fitur *polinomial*, optimasi

model, dan evaluasi performa model. Studi literatur ini bertujuan untuk memperkuat landasan teori dalam proses *training* dan pengujian model *machine learning* untuk mendeteksi kondisi pipa Normal atau Bocor.

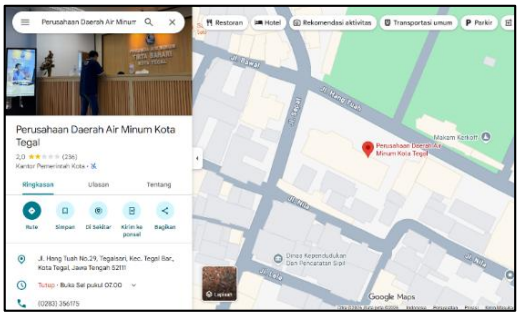
3.3. Waktu dan Tempat Penelitian

3.3.1. Waktu Penelitian

Pelaksanaan kegiatan penelitian terhadap objek ini dimulai pada 10 Desember 2025. Sejak tanggal tersebut, proses pengumpulan data, pengamatan, serta analisis yang berkaitan dengan objek penelitian mulai dilakukan secara bertahap dan terstruktur sesuai dengan rencana yang telah disusun sebelumnya.

3.3.2. Tempat Penelitian

Tempat pelaksanaan penelitian berlokasi di Perumda Tirta Bahari Kota Tegal. Pemilihan lokasi didasarkan pada kesesuaiannya dengan objek yang diteliti untuk menunjang efektivitas observasi dan pengumpulan data. Lokasi tempat penelitian ditunjukkan pada Gambar 3.3.



Gambar 3.3. Tempat Penelitian