

# Bab I

## Pendahuluan

### 1.1 Latar Belakang

Industri minyak dan gas saat ini sedang menghadapi tantangan yang kompleks, terutama insiden yang terjadi pada infrastruktur yang menyebabkan kerugian baik dalam hal keselamatan manusia maupun finansial. Insiden yang terjadi antara lain seperti korosi, kebocoran gas, dan karat. Kebocoran gas mempunyai dampak yang serius untuk lingkungan dan kesehatan. Kebocoran gas hidrokarbon berat yang terlepas ke lingkungan dapat mencemari udara sehingga dapat mempengaruhi indeks kebersihan udara secara global (Oktavianto et al., 2022). Selain itu, kebocoran pencemaran gas hidrokarbon dapat berpotensi menyebabkan penyakit dan kanker sehingga membahayakan para pekerja yang ada di industri perminyakan dan gas (Aldi, 2024).

Diperlukan adanya solusi untuk mengatasi masalah tersebut, dengan cara memantau transmisi gas menggunakan sistem yang dapat mendeteksi anomali pada sistem transmisi gas. Implementasi sistem dilakukan dengan cara menganalisis variabel yang dapat mendeteksi anomali data untuk dilakukan respons cepat menghindari bencana yang menimbulkan banyak kerugian. Pengamatan dan analisis data secara manual diperlukan waktu yang lama dan juga berisiko terjadinya kesalahan manusia. Penerapan *Supervised Learning* akan sangat bermanfaat dalam menganalisis data dan mendeteksi anomali pada transmisi gas (Aljameel et al., 2022).

Proses deteksi anomali data dilakukan dengan cara melakukan prosedur pengklasifikasian oleh model yang akan ditentukan di beberapa titik yang telah didapat pada data. Terdapat beberapa model yang cocok untuk diimplementasikan dengan baik berupa model *Supervised Learning* seperti *Feed Forward Neural Network* (FFNN) (Jayabharathi and Ilango, 2023), *Linear regression*, *Random Forest*, dan *Support Vector machine* (SVM). Selain itu, terdapat model pembelajaran mesin *Unsupervised Learning* seperti *Hierarchical Clustering* (Mulinka et al., 2020) dan *PCA*. Mengacu pada penelitian sebelumnya, penerapan mesin telah dilakukan berupa model *Isolation Forest* dan *One-Class SVM* untuk melakukan klasifikasi anomali data serta mengetahui evaluasi performa model dalam penerapan pada industri minyak dan gas tersebut (Lukito et al., 2023). Adapun penelitian yang menerapkan model *K-Nearest Neighbor* (KNN) dan *SVM* untuk deteksi anomali data yang melakukan klasifikasi untuk mengenali anomali pada data (Zuhri et al., 2024).

Pendekatan terbaru dalam deteksi anomali memanfaatkan algoritma *ma-*

*chine learning* untuk meningkatkan efisiensi dan akurasi deteksi pada sistem transmisi gas. Misalnya, metode *Isolation Forest* dan *One-Class SVM* telah terbukti efektif dalam mendeteksi deviasi ekstrem dari data normal pada operasi pipa gas, memberikan solusi yang andal untuk mencegah insiden lebih lanjut (Yan, 2020). Di sisi lain, model *Autoencoder* berbasis *Deep Learning* juga telah menunjukkan kemampuan yang tinggi dalam mengidentifikasi pola anomali pada data *multivariat*, menghasilkan deteksi yang lebih sensitif terhadap ketidaksesuaian dalam parameter operasional (Lee et al., 2020). Kombinasi pendekatan *supervised* dan *unsupervised* ini dapat digunakan untuk menciptakan sistem deteksi yang lebih holistik dan efisien di industri minyak dan gas.

Model berbasis *deep learning* seperti *Long Short-Term Memory* (LSTM) juga digunakan untuk mendeteksi anomali pada data *time-series* operasional pipa gas, menghasilkan akurasi yang tinggi dalam mengidentifikasi pola anomali (Rieza Fachrezi et al., 2024). Di sisi lain, teknik berbasis akustik digunakan untuk mendeteksi dan menentukan lokasi kebocoran gas dengan memanfaatkan analisis waktu-frekuensi untuk meningkatkan akurasi deteksi.

Sebagai bagian dari upaya meningkatkan akurasi deteksi anomali, penelitian ini juga menerapkan metode pelabelan awal *Manual Annotation* (MA) yang menggunakan pendekatan *Moving Average* untuk mengidentifikasi data normal dan anomali. MA dilakukan dengan menghitung nilai rata-rata bergerak dan deviasi standar dalam jendela waktu tertentu, yang digunakan sebagai batas toleransi untuk menentukan label data (Kim et al., 2023). Pendekatan ini memberikan fleksibilitas dalam pelabelan data tanpa memerlukan proses manual sepenuhnya, sekaligus mengurangi risiko bias manusia. Dalam konteks transmisi gas, MA memungkinkan identifikasi awal anomali berdasarkan pola perubahan data seperti tekanan atau aliran gas yang menyimpang dari norma operasional (Bamaqa et al., 2022).

Namun, untuk memastikan keandalan model deteksi anomali, diperlukan evaluasi yang komprehensif. Metrik seperti *Precision*, *Recall*, dan *F1-Score* digunakan untuk mengevaluasi akurasi deteksi dan kemampuan model dalam menemukan semua anomali, khususnya pada dataset yang tidak seimbang. Selain itu, *AUC-PR* dan *AUC-ROC* memberikan gambaran performa model dalam membedakan data normal dan anomali, bahkan pada distribusi data yang sangat tidak seimbang (Lukito et al., 2023).

Pada penelitian ini, dilakukan implementasi dengan tujuan mengetahui bagaimana cara penerapan model algoritma *Isolation Forest* dan *-Nearest Neighbor* (KNN) untuk mendeteksi anomali pada data transmisi gas. Data awal yang digunakan telah diberi label melalui metode *Manual Annotation*, sehingga model dalam penelitian ini berfungsi untuk melatih dan mengevaluasi performa deteksi anomali berdasarkan data berlabel tersebut. Maksud dari penelitian ini juga adalah untuk mengetahui uji evaluasi performa kinerja model serta menganalisis hasil penelitian dengan interpretasi manusia untuk menjelaskan output dari penelitian tersebut.

Pemilihan model *Isolation Forest* dan *K-Nearest Neighbor* (KNN) dalam penelitian ini didasarkan pada perbedaan pendekatan, keunggulan, dan rele-

vansi keduanya dalam mendeteksi anomali, sehingga memungkinkan analisis perbandingan yang mendalam. *Isolation Forest* dipilih karena keandalannya sebagai model *Unsupervised Learning* yang secara khusus dirancang untuk mendeteksi anomali melalui proses isolasi iteratif. Algoritma ini efisien untuk dataset besar, tetapi dalam penelitian ini, keberadaan label dari *Manual Annotation* digunakan untuk membimbing proses isolasi. Hal ini memberikan manfaat tambahan dalam meningkatkan akurasi deteksi pada data yang telah diberi label sebelumnya.

Sebaliknya, *K-Nearest Neighbor* (KNN) dipilih untuk memberikan perspektif yang berbeda karena pendekatan berbasis *Supervised Learning* yang mengklasifikasikan data berdasarkan kedekatannya dengan data lain dalam ruang multidimensi. Algoritma ini efektif dalam mendeteksi anomali yang lebih halus dan pola lokal pada dataset, seperti variasi data pipa yang tidak ekstrem tetapi tetap signifikan. Keunggulan KNN terletak pada fleksibilitasnya dalam menggunakan metrik jarak serta kesederhanaan implementasinya, meskipun memiliki keterbatasan dalam efisiensi untuk dataset besar.

Dengan membandingkan kedua model ini, penelitian bertujuan untuk mengevaluasi efektivitas *Isolation Forest* dan KNN berdasarkan karakteristik data transmisi gas, serta menganalisis kelebihan dan kekurangan masing-masing model. Hasil perbandingan ini diharapkan dapat memberikan panduan yang lebih jelas mengenai model mana yang paling sesuai untuk mendeteksi anomali pada sistem transmisi gas dengan mempertimbangkan efisiensi, akurasi, dan kemampuan interpretasi.

## 1.2 Rumusan Masalah

Dalam penelitian ini, terdapat beberapa masalah utama yang diidentifikasi terkait dengan deteksi anomali pada data operasional pipa minyak dan gas, yaitu:

1. Bagaimana merancang metode pelabelan awal menggunakan *Moving Average* yang efektif untuk deteksi anomali data operasional pipa minyak dan gas?
2. Bagaimana algoritma *Isolation Forest* dan *K-Nearest Neighbor* (KNN) dapat digunakan untuk mendeteksi anomali dengan mempertimbangkan keberadaan pelabelan awal?
3. Bagaimana mengoptimalkan hyperparameter pada masing-masing algoritma untuk optimasi performa deteksi anomali?
4. Bagaimana mengevaluasi hasil deteksi anomali menggunakan metrik evaluasi yang relevan?

## 1.3 Tujuan Penelitian

Penelitian ini bertujuan untuk:

1. Mengembangkan metode pelabelan awal berbasis *Moving Average* menggunakan *Moving Average* untuk deteksi anomali data operasional pipa minyak dan gas.

2. Menerapkan algoritma *Isolation Forest* dan *K-Nearest Neighbor* (KNN) dalam mendeteksi anomali pada data yang telah memiliki label awal.
3. Melakukan optimasi hyperparameter pada masing-masing algoritma agar lebih optimal dalam deteksi anomali.
4. Mengevaluasi performa deteksi anomali dari kedua algoritma dengan membandingkan hasil prediksi terhadap label awal menggunakan metrik yang relevan.

## 1.4 Batasan Penelitian

Penelitian ini memiliki beberapa batasan yang dirancang untuk memastikan fokus dan kelayakan penelitian, yaitu:

1. Data yang digunakan dalam penelitian ini berasal dari perusahaan Migas dengan menggunakan data operasional pipa minyak dan gas per jam dalam kurun waktu 1 tahun, dari bulan juli 2020 sampai agustus 2021. Pemilihan fitur berdasarkan analisis korelasi dan diambil 4 fitur dengan korelasi tertinggi.
2. Pelabelan awal menggunakan *Manual Annotation* berbasis *Moving Average* dengan parameter seperti ukuran jendela (*window size*) dan faktor toleransi ( $k$ ) yang ditentukan secara efektif dengan tuning parameter dan analisis pribadi berdasarkan visualisasi dan target rasio 5
3. Tahap pelabelan menggunakan metode *Majority Voting*
4. Evaluasi model dilakukan berdasarkan Metrik Evaluasi dan Visualisasi *ROC Curve*
5. Melakukan Analisis perbandingan cara kerja model *Isolation Forest* dan KNN dalam mendeteksi anomaly.

## 1.5 Manfaat Penelitian

### 1.5.1 Manfaat Akademik

1. **Pengembangan Metode Pelabelan Awal:** Memberikan kontribusi pada penelitian berbasis anomali dengan memperkenalkan metode pelabelan awal (*Manual Annotation*) berbasis *Moving Average*.
2. **Kontribusi ke Literatur Anomali:** Menambah wawasan tentang adaptasi algoritma *Isolation Forest* dan KNN untuk kasus yang memanfaatkan data berlabel.
3. **Analisis Algoritma:** Memberikan analisis perbandingan cara kerja model *Isolation Forest* dan knn dalam mendeteksi anomaly

### 1.5.2 Manfaat Praktis

1. **Peningkatan Keandalan Sistem Operasional:** Membantu perusahaan mengidentifikasi masalah operasional seperti kebocoran atau penyumbatan lebih dini, mengurangi risiko kerugian.
2. **Efisiensi Pemantauan:** Mengotomatisasi proses deteksi anomali untuk mengurangi ketergantungan pada inspeksi manual.
3. **Fleksibilitas dan Skalabilitas:** Pendekatan yang digunakan dapat diterapkan pada berbagai sistem operasional lainnya.

### 1.6 Rencana Kegiatan

Rencana kegiatan yang akan saya lakukan adalah sebagai berikut:

- **Literature Review:** Bertujuan untuk menentukan referensi atau landasan teori yang relevan dengan penelitian. Review dilakukan dengan meninjau jurnal, buku, dan penelitian sebelumnya terkait dengan deteksi anomali pada sistem transmisi gas. Fokus kajian meliputi algoritma *Isolation Forest*, *K-Nearest Neighbor* (KNN), dan metode pelabelan awal seperti *Manual Annotation* (MA) menggunakan *Moving Average*.
- **Pengumpulan dan Pengolahan Data:** Dilakukan pengumpulan dan pengolahan data untuk membangun model. Data yang akan digunakan sebagai bahan penelitian adalah data transmisi minyak dan gas. Data yang dikumpulkan berasal dari perusahaan minyak dan gas. Data ini mencakup kurun waktu 1 tahun, dari bulan Juli 2020 sampai Agustus 2021.

Pada proses pengolahan data akan dilakukan *preprocessing* untuk mempersiapkan sekaligus membersihkan data. Tahapan meliputi identifikasi nilai hilang. Dilanjutkan dengan *Exploratory Data Analysis* (EDA) untuk mengetahui karakteristik data melalui teknik visualisasi dan analisis korelasi antar fitur. Fitur yang relevan akan dipilih berdasarkan hasil analisis EDA.

- **Penerapan Model Pembelajaran Mesin:** Melakukan pelabelan dengan menggunakan *Moving Average* yang nantinya akan digunakan oleh model. lalu model deteksi anomali yang digunakan untuk analisis anomali pada data adalah *Isolation Forest* dan *K-Nearest Neighbor* (KNN), yang diimplementasikan menggunakan bahasa pemrograman Python dengan bantuan pustaka *scikit-learn*. Pelabelan awal menggunakan metode *Manual Annotation* dengan *Moving Average* akan digunakan untuk membimbing proses pelatihan dan evaluasi model.
- **Analisis dan Evaluasi Performa Model:** Hasil dari kedua model akan dianalisis dan dievaluasi untuk menentukan performa model terbaik yang dapat diimplementasikan pada deteksi anomali data. Analisis dilakukan menggunakan *Precision*, *Recall*, *F1-Score*, *AUC-PR*, dan

*AUC-ROC*. Perbandingan dilakukan untuk mengetahui kelebihan dan kelemahan dari masing-masing model.

- **Laporan Tugas Akhir:** Membuat laporan tugas akhir yang berisi keseluruhan proses kegiatan penelitian dari perencanaan sampai implementasi. Laporan mencakup latar belakang, metode, hasil penelitian, serta kesimpulan dan rekomendasi.

## 1.7 Jadwal Kegiatan

Laporan proposal ini akan dijadwalkan sesuai dengan tabel 1.1 berikut:

Tabel 1.1: Jadwal kegiatan proposal tugas akhir

| No | Kegiatan                             | Bulan ke- |  |  |  |   |  |  |  |   |  |  |  |   |  |  |  |   |  |  |  |   |  |  |  |
|----|--------------------------------------|-----------|--|--|--|---|--|--|--|---|--|--|--|---|--|--|--|---|--|--|--|---|--|--|--|
|    |                                      | 1         |  |  |  | 2 |  |  |  | 3 |  |  |  | 4 |  |  |  | 5 |  |  |  | 6 |  |  |  |
| 1  | Studi Literatur                      |           |  |  |  |   |  |  |  |   |  |  |  |   |  |  |  |   |  |  |  |   |  |  |  |
| 2  | Pengumpulan dan Pengolahan Data      |           |  |  |  |   |  |  |  |   |  |  |  |   |  |  |  |   |  |  |  |   |  |  |  |
| 3  | Penerapan Model Pembelajaran Mesin   |           |  |  |  |   |  |  |  |   |  |  |  |   |  |  |  |   |  |  |  |   |  |  |  |
| 4  | Analisis dan Evaluasi Performa Model |           |  |  |  |   |  |  |  |   |  |  |  |   |  |  |  |   |  |  |  |   |  |  |  |
| 5  | Laporan Tugas Akhir                  |           |  |  |  |   |  |  |  |   |  |  |  |   |  |  |  |   |  |  |  |   |  |  |  |