

1. Pendahuluan

Latar Belakang

Industri minyak dan gas merupakan salah satu sumber energi paling berharga dan dianggap sebagai indikator pertama pembangunan ekonomi. Secara global, permintaan minyak telah meningkat pesat karena digunakan di berbagai industri seperti industri transportasi dan rumah tangga. Sejalan dengan peningkatan permintaan terhadap industri minyak dan gas kita dihadapkan juga dengan ditemukannya berbagai masalah dan anomali yang dapat merusak dan merugikan pada saat pengolahan minyak dan gas yang akan berujung pada kecelakaan kerja yang akan merugikan keselamatan kerja maupun kesehatan. Beberapa contoh dari kasus anomali yang ada pada industri minyak dan gas adalah *oil spill*, korosi, kebocoran dan karat yang dalam hal ini sangat berbahaya dampaknya pada makhluk hidup dan lingkungan salah satunya pada penipisan lapisan ozon dan atau *global warming* [1]. Tumpahan minyak atau *oil spill* baik dari proses di kapal, pengeboran lepas pantai maupun akibat kecelakaan kapal. Polusi dari tumpahan minyak di laut merupakan sumber pencemaran laut yang selalu menjadi fokus perhatian dari masyarakat luas, karena akibatnya akan sangat cepat dirasakan oleh masyarakat sekitar pantai dan sangat signifikan merusak makhluk hidup di sekitar pantai tersebut.

Badan Dunia *Group of Expert on Scientific Aspects of Marine Pollution* (GESAMP) mencatat sekitar 6,44 juta ton per tahun masuk kandungan hidrokarbon ke dalam perairan laut dunia [2]. Dalam mengatasi masalah tersebut akan dilakukan prosedur klasifikasi data yang diterapkan dimana model nantinya akan menentukan secara spesifik dari keseluruhan data yang tersedia di beberapa titik pada dataset. Pada kasus ini terdapat beberapa model machine learning yang dapat digunakan yaitu menggunakan pembelajaran mesin terawasi (*Supervised Learning*) yaitu algoritma *Support Vector Machine (SVM)*, *K Nearest Neighbor (KNN)*, *Random Forest (RF)*, *Gradient Boosting (GB)*, dan *Decision Tree (DT)*[3] atau menggunakan pembelajaran mesin tidak terawasi (*Unsupervised Learning*) yaitu algoritma *Factor Analysis*, *PCA*, *Mixtures of Gaussians*, *ICA*, *Hidden Markov Models*, dan *StateSpace Models* [4].

Pada penelitian ini, dilakukan implementasi metode *K Nearest Neighbor (KNN)* untuk mendeteksi anomali pada data pipa gas. Anomali ini dapat ditemukan dalam satu atau lebih fitur yang saling terhubung. Penelitian ini juga dilakukan dengan tujuan untuk menganalisis perbandingan performa kinerja metode terbaik dengan menggunakan parameter yang tepat untuk mendukung hasil performa metode paling maksimal serta penelitian ini juga mengevaluasi hasil menggunakan interpretasi manusia apakah titik datanya terklasifikasi dalam anomali atau tidak. Pemilihan *KNN* sebagai metode deteksi anomali dalam industri minyak dan gas didasarkan pada kemampuannya untuk menangani data yang kompleks, fleksibilitas dalam berbagai jenis data, serta kesederhanaannya yang memungkinkan implementasi yang efektif. Dengan menggunakan *KNN*, perusahaan dapat meningkatkan keamanan operasional, mengurangi risiko kecelakaan, dan meningkatkan efisiensi produksi.

Topik dan Batasannya

Industri minyak dan gas merupakan sektor yang memegang peranan krusial dalam perekonomian global. Namun, industri ini juga dihadapkan pada sejumlah risiko signifikan, termasuk risiko kecelakaan kerja dan kegagalan peralatan. Insiden-insiden tersebut dapat mengakibatkan kerugian finansial yang besar serta membahayakan keselamatan pekerja dan lingkungan sekitar. Misalnya, kebocoran minyak yang tidak terdeteksi dapat mengakibatkan dampak ekologis yang serius, sementara kerusakan pada peralatan dapat mengganggu kelancaran proses produksi dan menimbulkan kerugian ekonomi yang substansial.

Dengan perkembangan teknologi yang pesat dan peningkatan kompleksitas operasi dalam industri minyak dan gas, terdapat kebutuhan yang semakin mendesak untuk mengembangkan sistem deteksi anomali yang akurat dan efisien. Anomali didefinisikan sebagai pola atau kejadian yang menyimpang dari data normal yang diharapkan. Deteksi anomali memainkan peran penting dalam mengidentifikasi kejadian tidak biasa dalam operasi sehari-hari, memungkinkan tindakan intervensi yang cepat untuk mencegah terjadinya insiden yang tidak diinginkan.

Berdasarkan latar belakang permasalahan yang sudah disajikan pada bagian sebelumnya, berikut adalah rumusan masalah yang disajikan:

1. Bagaimana implementasi algoritma *K Nearest Neighbor (KNN)* untuk melakukan klasifikasi anomali pada data operasional pipa gas?
2. Bagaimana implementasi *ROC Curve* untuk evaluasi dalam menentukan metode pembelajaran mesin *K Nearest Neighbor (KNN)* sebagai metode yang efektif dan memiliki performa yang bagus?

Batasan-batasan ini ditetapkan untuk memastikan bahwa penelitian dapat diselesaikan dalam jangka waktu yang ditetapkan dan dengan sumber daya yang tersedia. Pada penelitian selanjutnya, batasan-batasan ini dapat dihilangkan atau dilonggarkan untuk mengeksplorasi potensi yang lebih luas dari deteksi anomali dalam industri minyak dan gas.

Tujuan

Berdasarkan rumusan masalah yang dijelaskan pada bagian sebelumnya, maka penelitian ini akan memiliki beberapa tujuan diantaranya adalah:

1. Melakukan implementasi algoritma *K Nearest Neighbor (KNN)* dalam melakukan klasifikasi dan mendeteksi anomali data operasional pipa gas.
2. Melakukan evaluasi metode *KNN* menggunakan *ROC Curve*.

Organisasi Tulisan

Rencana penelitian akan mencakup beberapa proses yang dimulai dengan melakukan kajian pustaka. Kajian pustaka dilakukan dengan tujuan untuk menentukan referensi dan acuan ilmiah pada penelitian ini dengan penelitian yang sudah pernah dilakukan sebelumnya. Setelah melakukan kajian pustaka sebagai mendukung landasan teori penelitian akan dilakukan pengumpulan dan pengolahan data model pembelajaran mesin.

Dalam membangun model pembelajaran mesin dibutuhkan aset data sebagai masukkan dalam model pembelajaran mesin. Proses pengambilan data ini tidak dapat dilewatkan karena dalam penelitian ini sangat membutuhkan sumber data sebagai pilar dalam menerapkan implementasi model pembelajaran mesin pada penelitian ini. Data yang dikumpulkan sebagai masukkan dalam model pembelajaran mesin merupakan data operasional pipa gas alam.

Data merupakan data operasional dari kurun waktu dua tahun terakhir, namun dalam lingkup proses implementasi pada penelitian ini, model hanya akan menggunakan dua sumber data saja yaitu data dari rentang bulan Juli 2020. Setelah itu akan dilakukan pelabelan anomali menggunakan metode *moving average* untuk memprediksi titik anomali yang ada pada *dataset* lalu membandingkannya dengan titik anomali data yang sebenarnya menggunakan metode *KNN* dan setelah dilakukan pengujian berbagai nilai sebagai parameter *KNN*. Setelah dilakukan berbagai pengujian parameter *KNN*, akan dilakukan penilaian dengan menggunakan *ROC Curve* untuk mengetahui seberapa bagus kinerja *KNN* untuk mendeteksi anomali.