

BAB I

PENDAHULUAN

1.1 Latar Belakang

Revolusi industri 4.0 memperkenalkan sistem industri berbasis *smart factory* dengan sistem otomatisasi yang berintegrasi sistem fisik dan sistem digital. Industri yang berbasis *smart manufacturing* pada industri 4.0 membutuhkan solusi dalam hal kesehatan mesin industri. Solusi yang ditawarkan yaitu *predictive maintenance* (Çinar dkk, 2020). *Machine Learning* (ML) adalah salah satu pengaplikasian yang populer digunakan dalam *predictive maintenance*. ML merupakan program pembelajaran komputer yang memiliki kemampuan mengolah data dan informasi data yang telah diinputkan. ML dapat membantu memecahkan kesulitan dalam *vision*, *big data*, robotika, dan pengenalan suara (Theissler dkk, 2021). Data yang digunakan sebagai input ML adalah data yang ditangkap oleh sensor yang terletak di sekitar peralatan, mesin, atau komponen (seperti tekanan, getaran, suhu, viskositas, akustik). Data yang terkumpul dapat digunakan untuk mendeteksi anomali pada mesin berbasis ML (Çinar dkk, 2020). Beberapa metode yang bisa digunakan berbasis ML, seperti *neural work*, *fuzzy logic*, *genetic algorithm*, dan *support vector machine* (SVM) (Fathurrohman dkk, 2019).

Dataset yang digunakan dalam ML, sebaiknya memiliki kelas yang seimbang. Ketidakseimbangan kelas pada dataset biasanya akan memiliki bias dengan menghasilkan tingkat akurasi yang tinggi pada data kelas mayoritas dan menghasilkan tingkat akurasi rendah pada data kelas minoritas. Pada umumnya dalam pengaplikasian ML ditemukan kumpulan dataset dalam berbagai tingkat ketidakseimbangan mulai ketidakseimbangan kecil, sedang sampai ekstrim. (Erllin dkk, 2022). Adapun teknik-teknik untuk memodifikasi kelas mayoritas dan kelas data minoritas secara luas yaitu *undersampling* dan *oversampling*. Pada teknik *undersampling* penyeimbangan data dilakukan dengan cara menghapus atau membuang sejumlah data pada kelas mayoritas, yang dapat mengakibatkan hilangnya informasi. Pada teknik *oversampling* penyeimbangan data dilakukan

dengan cara meningkatkan sejumlah data pada kelas minoritas (Erlin dkk, 2022). Adapun teknik-teknik yang lain seperti *random undersampling* dan *random oversampling*. Pada teknik *random undersampling* penyeimbangan data dilakukan dengan cara menghapus atau mengeliminasi secara acak sejumlah data pada kelas mayoritas. Pada teknik *random oversampling* penyeimbangan data dilakukan dengan meningkatkan sejumlah data secara acak pada kelas minoritas (Amelia dkk, 2021). Menurut (Mohammed dkk, 2020) melakukan eksperimen menggunakan dua metode teknik yaitu *undersampling* dan *oversampling*. Berdasarkan hasil penelitian tersebut teknik *oversampling* berkinerja lebih baik daripada teknik *undersampling* untuk pengklasifikasian dan perolehan skor di berbagai matrix evaluasi (Mohammed dkk, 2020). Dalam penelitian tersebut, dataset yang digunakan mengalami imbalance data (Mohammed dkk, 2020). *Synthetic Minority Oversampling Technique* (SMOTE) adalah salah satu metode *oversampling* yang populer digunakan untuk menyeimbangkan data antar kelas dalam ML. Keunggulan dari metode SMOTE adalah dapat menangani permasalahan imbalance data dengan prinsip menambah sejumlah data kelas minoritas supaya setara dengan kelas mayoritas dengan membangkitkan data buatan (Erlin dkk, 2022).

Pada penelitian yang dilakukan oleh Erlin dkk (2022), SMOTE digunakan untuk memprediksi kesehatan jantung. Dataset yang digunakan penelitian tersebut berjumlah 299 data yang dibagi 2 bagian, 239 data digunakan training (80%) dan 60 data digunakan untuk testing (20%). Dataset pada penelitian tersebut mempunyai 13 variabel, yang terdiri dari umur, penurunan sel darah merah, tingkat enzim dalam darah, diabetes, ejection fraction, tekanan darah, trombosit dalam darah, tingkat kreatinin, tingkat sodium, jenis kelamin, perokok, waktu, dan peristiwa kematian. Berdasarkan pada hasil penelitian tersebut terjadi peningkatan nilai skor akurasi, presisi, dan *sensitivity* yaitu 0,87 (87%), 0,83 (83%), dan 0,85 (85%) ketika sebelum dilakukan SMOTE menjadi 0,90 (90%), 0,90 (90%), dan 0,91 (91%) ketika SMOTE digunakan. Dengan hasil tersebut membuktikan bahwa dampak penambahan SMOTE yang dilakukan dapat mengatasi masalah *overfitting* (Erlin dkk, 2022).

Pada penelitian yang dilakukan oleh Dutta dkk (2021), Integrasi SMOTE dengan metode klasifikasi ML diterapkan untuk memprediksi dan mendeteksi sindrom ovarium polikistik. Dataset yang digunakan penelitian tersebut mengalami *imbalance* data. Data dalam penelitian tersebut dengan jumlah 541 data terbagi dalam dua kelas data yaitu dengan berjumlah 178 kelas data positif dan berjumlah 363 kelas data negatif. Penelitian tersebut menggunakan metode performansi *random forest*, *decision tree*, *support vector machine* (SVM), dan *k-nearest neighbor* (KNN). Berdasarkan hasil penelitian tersebut untuk rata-rata tingkat akurasi metode ML dengan *random forest* 89%, *decision tree* 89%, SVM 91%, dan KNN 90%. Berdasarkan hasil penelitian tersebut metode performansi SVM dengan 91% memiliki tingkat akurasi paling tinggi dibandingkan dengan metode ML yang lain (Dutta dkk, 2021).

Salah satu metode ML yang digunakan untuk klasifikasi adalah *support vector machine* (SVM). Dalam pemantauan dan analisa kondisi mesin (*condition monitoring*), metode SVM dapat digunakan untuk mengenali pola khusus dari sinyal akustik mesin yang didapatkan. Lalu sinyal akustik tersebut membentuk pola-pola yang akan diklasifikasikan sesuai dengan permasalahan mesin (Widodo & Yang, 2007). Metode klasifikasi SVM dalam hal model lebih baik daripada metode *neural network* berdasarkan hasil performansi klasifikasi (Fathurrohman dkk, 2019). Maka dari itu, metode SVM telah berhasil diterapkan di berbagai macam aplikasi seperti deteksi wajah, verifikasi, arahan teks, klasifikasi kondisi mesin, dan lain sebagainya karena memiliki tingkat keakurasian yang cukup tinggi (Widodo & Yang, 2007).

Dalam penelitian yang dilakukan oleh Dzilfadhilah pada tahun (2021) telah melakukan klasifikasi kerusakan mesin pompa berdasarkan SVM dengan menggunakan data akustik. Objek penelitian tersebut adalah mesin pompa yang sengaja dimodifikasi dengan kerusakan *missalignment*, *unbalance*, dan *bearing fault*. Penelitian ini menggunakan fitur akustik yakni *root mean square*, *kurtosis*, *skewness*, *crest factor*, *shape factor*, dan *instaneous frequency*. Jumlah data yang digunakan dalam pembelajaran mesin (*training data*) sebesar 50% dari total jumlah

dataset dan jumlah data yang digunakan dalam pengujian model yang dihasilkan 50% dari total jumlah dataset. Berdasarkan hasil penelitian tersebut tingkat akurasi dari klasifikasi kerusakan mesin pompa dengan menggunakan metode SVM Linear adalah sebesar 96,7%, metode SVM Quadratic sebesar 100%, dan metode SVM Cubic sebesar 96,7%. Dari beberapa metode yang telah digunakan pada penelitian tersebut metode SVM memiliki tingkat akurasi paling tinggi dibandingkan dengan metode *decision tree* yang menghasilkan tingkat akurasi sebesar 80% dan metode KNN sebesar 76,7% (Dzilfadhilah, 2021).

Aciandra pada tahun (2022) telah melakukan penelitian deteksi anomali pada mesin pompa berdasarkan fitur akustik dan ML dengan menggunakan data akustik. Variasi *signal to noise ratio* (SNR) yang digunakan pada penelitian tersebut adalah -6 dB, 0 dB, 6 dB. Data yang digunakan pada penelitian tersebut yang semula tidak seimbang dengan jumlah 702 data pada kelas normal dan 100 data pada kelas abnormal mesin pompa. Penelitian tersebut menggunakan metode *undersampling* dengan penyeimbangan pada data dengan cara mengurangi atau membuang sejumlah data pada kelas normal. Fitur akustik yang digunakan adalah 5 fitur terbaik yaitu *root mean square*, *standard deviation*, *band power*, *skewness*, dan SNR. Berdasarkan hasil penelitian tersebut didapatkan beberapa hasil performansi model deteksi anomaly, pada variasi SNR -6 dB yang tingkat rata-rata akurasi mencapai 79,76 %, pada variasi SNR 0 dB yang tingkat rata-rata akurasi mencapai 88,7%, pada variasi SNR 6 dB yang tingkat rata-rata akurasi mencapai 63,81% dari model SVM (Aciandra, 2022).

Penelitian yang diusulkan ini merupakan penelitian terdahulu yang dilakukan oleh Tiffani Febbiola Aciandra yang berjudul “Deteksi Anomali Pada Mesin Pompa Berdasarkan Fitur Akustik dan *Machine Learning*”. Pada penelitian sebelumnya, penggunaan teknik untuk menyeimbangkan kelas mayoritas dan minoritas menggunakan teknik *undersampling* dimana teknik tersebut memiliki kelemahan yaitu dapat mengakibatkan hilangnya informasi penting pada kelas mayoritas. Dalam hal mengatasi terjadinya ketidakseimbangan data, metode yang digunakan penelitian ini adalah teknik SMOTE. Total dataset setelah proses

SMOTE adalah 1404 data. Dataset yang digunakan nilai SNR 6 dB. Pada penelitian ini dilakukan proses klasifikasi dengan SVM. Penelitian ini diharapkan dapat memberikan referensi baru dalam peingintegrasian SMOTE dan SVM untuk deteksi anomali pada mesin pompa.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, perumusan masalah yang didapatkan adalah bagaimana peromansi model dari hasil pengintegrasian *synthetic minority oversampling technique* (SMOTE) dan *support vector machine* (SVM) untuk deteksi anomali pada mesin pompa?

1.3 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah untuk mengetahui peromansi model dari hasil pengintegrasian *synthetic minority oversampling technique* (SMOTE) dan *support vector machine* (SVM) untuk deteksi anomali pada mesin pompa.

1.4 Manfaat Penelitian

Adapun manfaat pada penelitian ini adalah sebagai berikut :

- 1) Menambah pengetahuan baru bagi siapapun yang ingin mengetahui bidang ilmu *machine learning*
- 2) Memberikan pengembangan baru dalam analisa sistem pemeliharaan prediktif

1.5 Batasan Penelitian

Adapun batasan pada penelitian ini adalah sebagai berikut :

- 1) Penggunaan dataset berupa *wavfile* dari mesin pompa industri dengan nilai SNR 6dB
- 2) Ekstraksi fitur dilakukan dalam domain waktu dan domain frekuensi

1.6 Hipotesa Penelitian

Adapun hipotesa awal dalam penelitian ini adalah sebagai berikut:

- 1) Data yang digunakan berdistribusi normal dan acak
- 2) Seleksi terhadap fitur akustik yang digunakan untuk membuat model didasarkan pada karakteristik data

