

Deteksi Kebocoran Pipa Air Menggunakan *Machine Learning* dengan Jaringan Nirkabel IEEE 802.15.4

Kurniawan Saputra, M. Udin Harun Al Rasyid*, Muh. Zen Samsono Hadi
Politeknik Elektronika Negeri Surabaya (PENS)
Jl. Raya ITS, Kampus PENS Sukolilo, Surabaya, Indonesia
Email: putr4.iwan@gmail.com, udinharun@pens.ac.id, zenhadi@pens.ac.id
*Corresponding author: M. Udin Harun Al Rasyid

Abstrack - Pipes are the most economical and safest way to distribute products such as water, petrochemicals, gases and other liquids. The use of parameters such as flow rate and balance to determine leaks in pipes is a relatively simple and affordable method. There are some limitations to using this method, such as determining the size and location of the leak. One of the best ways is to train the system using machine learning algorithm models such as Nave Baiyes, Multi Layer Perceptron (MLP) and Support Vector Machine (SVM). The algorithm model is trained using two dataset models, namely raw dataset and process dataset, z-score data normalization method, and utilization of kernel functions. The results of the leakage size classification with the best accuracy were obtained using a raw dataset with a polynomial kernel SVM algorithm model of 98.25% with 99.1% recall, 99.8% precision, and 99.5% F-measure. In the classification of the location of the leak, the raw dataset has good accuracy with the MLP algorithm model of 89.31%, while the SVM kernel polynomial accuracy value is 88.96%, the difference of 0.35% can be considered good enough to pursue its accuracy. The model that is formed is quite good at classifying the size and location of leaks, so it is quite important to select dataset model, normalize data and optimize the use of kernel functions.

Keywords – Pipe leakage, Volume Balance, Support Vector Machine

Intisari – Pipa adalah cara paling ekonomis dan paling aman dalam mendistribusikan hasil produk seperti air, petrokimia, gas, dan cairan lainnya. Penggunaan parameter seperti debit aliran dengan memperhatikan kesetimbangannya merupakan metode yang cukup mudah dan murah untuk menentukan kebocoran pada pipa. Ada beberapa keterbatasan penggunaan metode ini, seperti menentukan ukuran dan lokasi kebocoran. Salah satu cara yang tepat adalah melatih sistem menggunakan model algoritma machine learning seperti *naïve baiyes*, *Multi Layer Perceptron (MLP)* dan *Support Vector Machine (SVM)*. Model algoritma dilatih menggunakan dua model *dataset* yaitu *raw dataset* dan *process dataset*, metode z-score normalisasi data, dan pemanfaatan fungsi kernel. Hasil klasifikasi ukuran kebocoran dengan akurasi terbaik diperoleh menggunakan *raw dataset* dengan model algoritma SVM kernel *polynomial* sebesar 98,25% dengan *recall* 99,1%, presisi 99,8%, dan *F-measure* 99,5%. Pada klasifikasi lokasi kebocoran *raw dataset* memiliki akurasi yang baik dengan model algoritma MLP sebesar 89,31%, sedangkan SVM kernel *polynomial* nilai akurasi sebesar 88,96%, selisih 0.35% dapat dinilai cukup baik untuk mengejar keakurasiannya. Model yang terbentuk cukup baik dalam mengklasifikasikan ukuran dan lokasi kebocoran, sehingga cukup penting pula pemilihan model *dataset*, menormalisasi data dan optimalisasi penggunaan fungsi kernel.

Kata Kunci – Kebocoran pipa, Kesetimbangan Volume, *Support Vector Machine*.

I. PENDAHULUAN

Dalam industri pengolahan air minum, minyak dan gas bumi, maupun petrokimia salah satu metode pendistribusian hasil produk adalah melalui jaringan pipa. Distribusi hasil produk melalui pipa ini dinilai paling ekonomis dan aman. Jaringan pipa adalah jalur kehidupan yang sangat penting untuk hampir setiap aktivitas kehidupan modern. Terlepas dari manfaat ini,

jaringan pipa juga membawa permasalahan yang umum yaitu potensi kebocoran pipa. Hal ini dapat diakibatkan oleh beberapa faktor diantaranya adalah kegagalan material pipa, sifat media yang korosif yang diakibatkan oleh tanah maupun udara, suhu, tekanan lingkungan, kegagalan pada saat pemasangan pipa, akibat perubahan struktur geologi dan hal lainnya yang diakibatkan oleh kegagalan lain yaitu akibat kemungkinan aksi sabotase yang dilakukan oleh pihak – pihak yang tidak bertanggung jawab. Dalam pelaksanaannya deteksi kebocoran masih banyak dilakukan menggunakan metode konvensional inspeksi lapangan secara langsung, seperti teknik sounding. Metode tersebut dinilai membutuhkan waktu yang cukup lama, peralatan yang kompleks dan terbatasnya tenaga yang berpengalaman.

Adanya permasalahan tersebut dibutuhkan sistem yang dapat memantau secara *realtime* dengan memanfaatkan parameter-parameter yang tersedia seperti tekanan, laju alir, dan suhu pipa. Akan tetapi data-data tersebut perlu dilakukannya analisa sehingga dapat mengidentifikasi ukuran dan lokasi kebocoran secara cepat dan tepat. Dalam beberapa penelitian yang telah dilakukan untuk mendeteksi kebocoran pipa menggunakan algoritma *Support Vector Machine* (SVM) [1-2]. Ada juga beberapa peneliti yang menggunakan *Neural Network* dalam mendeteksi kebocoran pipa [3]. Penelitian selanjutnya menggunakan data *time series* tekanan dan volume air dan metode regresi sebagai algoritmanya [4], ada pula peneliti dalam deteksi kebocoran pipa dan lokasinya menggunakan metode berbasis *fuzzy* [5], maupun penggunaan metode *K-Nearest Neighbor* [6]. Beberapa penelitian lainnya yang telah dikembangkan, Y. Liu, dkk, menggunakan metode SVM dan Fitur Frekuensi Waktu untuk melakukan analisa sinyal yang dihasilkan oleh beberapa node sensor [7]. Penelitian lain juga melakukan perbandingan dengan beberapa metode menggunakan deviasi hasil pengukuran tekanan dan laju aliran sebagai data yang akan diolah kedalam metode algoritma *Bayesian Probabilistic Analysis*, SVM, dan *Artificial Neural Network* (ANN) [8].

Metode kesetimbangan aliran fluida antara *inlet* dan *outlet* pipa juga dilakukan sehingga dapat dideteksi kebocoran pipa [9]. Selain itu adanya solusi baru dalam pengiriman data suatu node sensor dengan menggunakan protokol komunikasi LoRaWAN® untuk mendapatkan data dari *flow* meter dan temperatur sensor, kemudian diolah menggunakan regresi logistic dan fungsi sigmoid untuk mengestimasi probabilitas dari kebocoran pipa [10]. Penelitian lain membandingkan performa model algoritma untuk mendeteksi kebocoran pada pipa secara tepat dan real time. Model algoritma yang digunakan adalah *Random Forest*, *Decision Tree*, *Neural Network*, dan SVM [11]. Ada pula penelitian lain yang mendiagnosa kebocoran pipa berdasarkan informasi yang diperoleh dengan mensimulasikan perilaku dari fluida pada simulasi pilot, data yang digunakan adalah laju alir pada *inlet* dan *outlet* pipa, yang kemudian dilatih menggunakan *Multilayer Perceptron Neural Network* [12].

Oleh karena itu, dalam upaya mengurangi kegagalan dalam sistem distribusi produk terutama pada kegagalan pipa, maka dibutuhkan suatu model sistem yang dapat mendeteksi kebocoran pipa secara tepat. Model atau sistem algoritma deteksi yang efektif adalah model yang dapat memprediksi secara tepat dan respon yang cepat dalam memprediksi kebocoran pada pipa. Selain pemilihan format *dataset*, proses *pre-processing* dan optimalisasi penggunaan fungsi kernel sangat menentukan, karena akan mempengaruhi hasil prediksi pada sistem yang dibangun. Penelitian ini mengusulkan desain perangkat, model algoritma *machine learning* dan pengembangan sistem deteksi kebocoran pipa model dan format data yang tepat sehingga model algoritma yang dilatih dapat memprediksi besar ukuran dan lokasi kebocoran pada pipa.

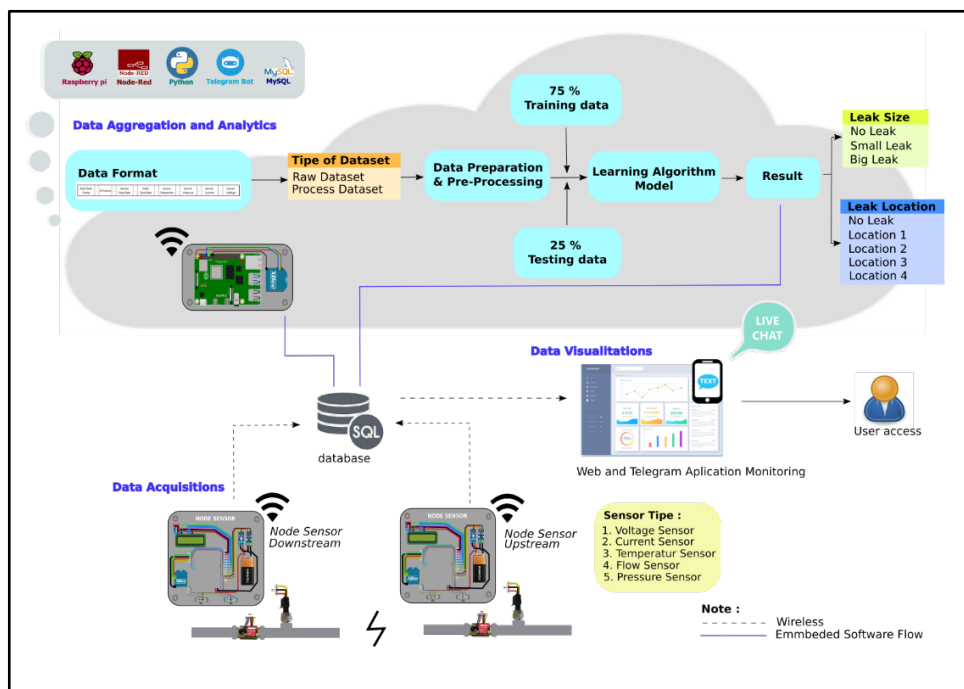
Artikel ini disusun sebagai berikut. Bagian 2 menjelaskan desain sistem yang diusulkan dan metode penelitian menggunakan metode kesetimbangan dan *machine learning* dalam penelitian ini. Bagian 3 menjelaskan eksperimen dan analisis yang berisi pengujian perangkat keras sistem sensor deteksi kebocoran, perolehan *dataset*, *data processing*, perbandingan beberapa model algoritma *machine learning*, prediksi ukuran dan lokasi kebocoran. Akhirnya, kesimpulan dan rencana untuk pekerjaan kami di masa depan disediakan di Bagian 4.

II. SIGNIFIKANSI STUDI

Berikut ini penjelasan signifikansi studi dari penelitian yang terdiri desain sistem dan metode penelitian yang digunakan untuk mendapatkan sebuah sistem deteksi ukuran dan lokasi kebocoran pipa.

A. Desain Sistem

Pada studi ini, penulis mengusulkan sistem deteksi kebocoran untuk menyelesaikan permasalahan ukuran dan lokasi kebocoran. Pengujian dan pengumpulan *dataset* kebocoran dibuat melalui model jaringan pipa skala laboratorium. Jaringan pipa yang dibuat menggunakan pipa *PolyVinyl Chloride* dengan ukuran $\frac{1}{2}$ inch. Informasi kelima parameter dikirim oleh masing-masing node sensor, kemudian diterima oleh *gateway* menggunakan protokol IEEE 802.15.4. *Gateway* terdiri dari XBee S2C tipe *Coordinator* yang terintegrasi dengan *Raspberry Pi* yang memiliki fungsi untuk melakukan penyimpanan dan pengolahan data sehingga dapat memprediksi ukuran dan lokasi kebocoran. Sistem dapat mengklasifikasikan tiga jenis ukuran kebocoran yaitu, *no leak*, *small leak* dan *big leak*. Selain ukuran, lokasi kebocoran diklasifikasikan kedalam empat klasifikasi *no leak*, *location 1*, *location 2*, *location 3*, dan *location 4*.

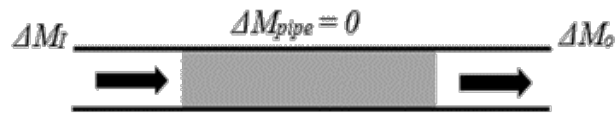


Gambar 1. Desain Sistem

Gambar 1 menunjukkan desain sistem, *gateway* akan menyimpan pada *database* dan memproses data. Ada beberapa *service* aplikasi yang ditanamkan kedalam *single board computing* diantaranya 1). Platform IoT yang memiliki fungsi peran sebagai pengolah data, *dashboard*, *generator*, dan *command line tool*, 2). *Phyton* yang berfungsi sebagai model klasifikasi sehingga menghasilkan informasi kebocoran.

B. Kebocoran Pipa dengan Pendekatan Metode Kesetimbangan

Penelitian ini memperhatikan kesetimbangan material yang masuk dan keluar sistem. Metode ini terbatas pada pendekatan dasar seperti penyeimbang massa dimana massa yang masuk ΔM_I dalam interval satuan waktu Δt dan keluar ke massa ΔM_O dalam interval satuan waktu [13].



Gambar 2. Kestimbangan massa pada pipa dalam kondisi steady state.

Kebocoran massa dihitung menggunakan persamaan nomor 1 sebagai berikut,

$$\Delta M_{Leak} = \Delta M_I - \Delta M_O \tag{1}$$

Karena pengaruh suhu sangat kecil terhadap perubahan masa jenis air [14]. Maka didapatkan kesetimbangan volume antara volume yang masuk ΔV_I dengan volume yang keluar ΔV_o menggunakan persamaan nomor 2,

$$\Delta V_{leak} = \Delta V_I - \Delta V_o \tag{2}$$

C. *support vector machine (SVM)*

Pengklasifikasian menggunakan metode algoritma SVM diterapkan karena kemampuan dan reputasi generalisasi yang luar biasa dalam melatih *dataset* sehingga mendapatkan akurasi yang tinggi [15], [16]. SVM memperkenalkan fungsi kernel, $K(X_n, X_i)$, yang mengubah ruang data asli menjadi ruang baru dengan dimensi yang lebih tinggi; proses ini mencakup fungsi transformasi dengan produk titik $\phi(X)$ (persamaan nomor 3). Dengan demikian fungsi hyperplane dapat ditulis dalam persamaan nomor 4.

$$K(X_n, X_i) = \phi(X_n)\phi(X_i) \tag{3}$$

$$f(X_i) = \sum_{n=1}^N \alpha_n y_n K(X_n, X_i) + b \tag{4}$$

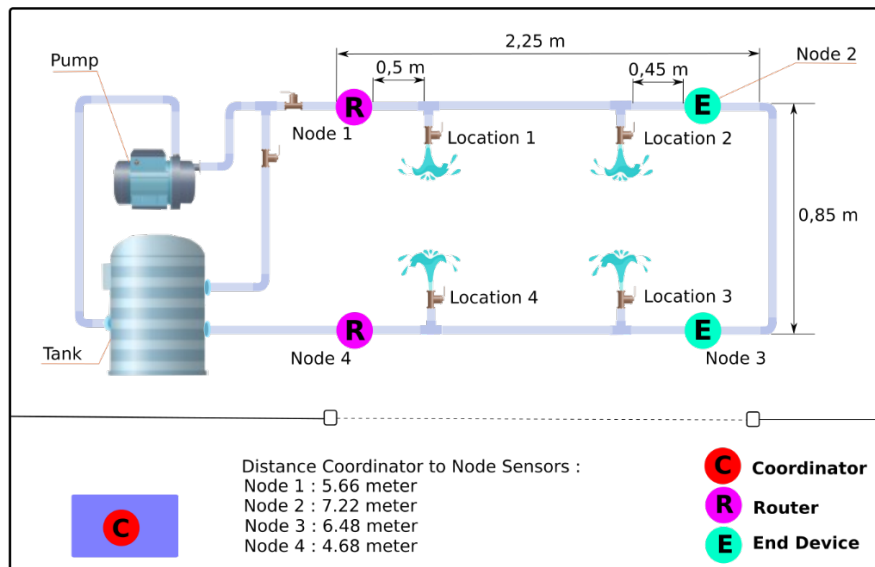
dimana, X_n adalah data vektor pendukung, n adalah pengali lagrange dan y_n adalah label kelas keanggotaan (+1, -1) dengan $n = 1, 2, 3, \dots, N$. Selain beberapa metode algoritma seperti *Naïve Bayes* dan *Multi Layer Perceptron (MLP)*. Tabel I menunjukkan perbandingan empat fungsi kernel pada algoritma SVM, yaitu *linear*, *radial basis function (RBF)*, *sigmoid*, dan *polinomial*.

TABEL I
FUNGSI KERNEL

No	Fungsi Kernel	Formula	Parameter Optimasi
1	Linear	$K(X_n, X_i) = (X_n, X_i)$	C dan γ
2	RBF	$K(X_n, X_i) = \exp(-\gamma \ X_n - X_i\ ^2) + C$	C dan γ
3	Sigmoid	$K(X_n, X_i) = \tanh(\gamma(X_n, X_i) + r)$	C, γ dan r
4	Polynomial	$K(X_n, X_i) = (\gamma(X_n, X_i) + r)^d$	C, γ, r dan d

1) *Pengambilan data untuk proses learning*

Untuk mendapatkan *dataset*, penulis membangun sebuah model jaringan pipa skala laboratorium. Jaringan pipa berbahan *PolyVinyl Chloride* disusun berbentuk ‘U’. Pada Gambar 3 menunjukkan dimensi model jaringan pipa dan jarak antar node sensor dan *gateway*.



Gambar 3. Model Jaringan Pipa Skala Laboratorium

Besar ukuran kebocoran dilakukan sebanyak 9 titik bukaan *valve*, adapun pada bukaan *valve* sebesar 10° sampai dengan 90°. Tabel II menunjukkan label *class* yang menggambarkan besar ukuran kebocoran pada pipa. Sedangkan untuk lokasi kebocoran ditandai dengan lokasi dari titik kebocoran.

TABEL II
CLASS LABEL UKURAN KEBOCORAN

Output	Class	Dekripsi
0	No Leak	No Leak atau <i>valve</i> tertutup
1	Small Leak	Bukaan <i>valve</i> sebesar 10°, 20°, and 30°
2	Big Leak	Bukaan <i>valve</i> sebesar 40°, 50°, 60°, 70°, 80°, and 90°

Data yang terkumpul untuk *raw dataset* sebanyak 13.239 data dengan 9 fitur data, dan *process dataset* sebanyak 52.956 data dan 7 fitur data. Tabel III menunjukkan struktur *process dataset* yang terdiri dari beberapa fitur.

TABEL III
FITUR PROCESS DATASET

Fitur	Deskripsi
<i>SensorID_i</i>	ID Setiap Sensor
<i>Value_i</i>	Nilai hasil pengukuran sensor
<i>Average_i</i>	Rata-rata 5 data pengukuran sensor
<i>Diff_ref_i</i>	Nilai kesetimbangan volume terhadap sensor 1
<i>Diff_sen_i</i>	Nilai kesetimbangan volume dari sensor sebelumnya
<i>Temp_{sensor_i}</i>	Data pengukuran suhu sensor
<i>Class</i>	Klasifikasi besar ukuran kebocoran
<i>Size</i>	Derajat bukaan <i>valve</i> kebocoran
<i>Leak_spot</i>	Titik lokasi kebocoran

Representasi kedua *dataset* untuk *raw dataset* dan *process dataset* dapat terlihat pada Tabel IV. Untuk *raw dataset* menampilkan informasi data langsung hasil pengukuran dan *process dataset* memiliki fitur yang merujuk pada penelitian [17].

TABEL IV
DATASET (A). RAW DATASET, (B). PROCESS DATASET

flow_sensor_1	temp_sensor_1 ...	flow_sensor_4	temp_sensor_4	Class	Size	Location	Leak_spot
0.00	27.97 ...	0.00	26.92	NoLeak	0	No	No
0.00	27.92 ...	0.00	26.90	NoLeak	0	No	No
0.00	27.92 ...	0.00	26.78	NoLeak	0	No	No
0.00	27.82 ...	0.00	26.86	NoLeak	0	No	No
0.00	28.12 ...	0.00	26.76	NoLeak	0	No	No
...
19.77	27.02 ...	18.84	26.55	SmallLeak	10	Line1	Location_1
19.68	26.90 ...	18.73	26.37	SmallLeak	10	Line1	Location_1
19.81	26.69 ...	18.75	26.55	SmallLeak	10	Line1	Location_1
19.77	26.67 ...	18.77	26.49	SmallLeak	10	Line1	Location_1
19.68	26.98 ...	18.73	26.43	SmallLeak	10	Line1	Location_1

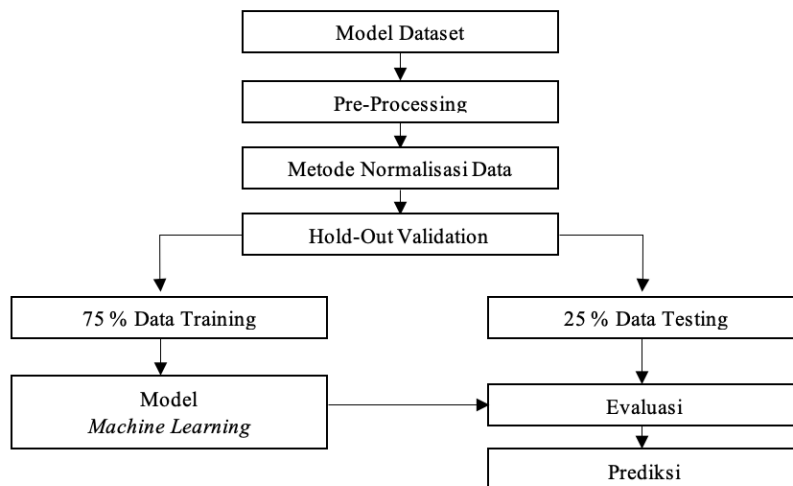
(a)

sensorID	flow_sensor_i	average_sens_i	diff_ref_i	diff_sen_i	temp_sensor_i	Class	Size	Location	Leak_spot
1001	0.00	0.00	0.0	0.0	27.97	NoLeak	0	No	No
1001	0.00	0.00	0.0	0.0	27.92	NoLeak	0	No	No
1001	0.00	0.00	0.0	0.0	27.92	NoLeak	0	No	No
1001	0.00	0.00	0.0	0.0	27.82	NoLeak	0	No	No
1001	0.00	0.00	0.0	0.0	28.12	NoLeak	0	No	No
...
1001	19.77	19.76	0.0	0.0	27.02	SmallLeak	10	Line_1	Location_1
1001	19.68	19.74	0.0	0.0	26.90	SmallLeak	10	Line_1	Location_1
1001	19.81	19.75	0.0	0.0	26.69	SmallLeak	10	Line_1	Location_1
1001	19.77	19.75	0.0	0.0	26.67	SmallLeak	10	Line_1	Location_1
1001	19.68	19.74	0.0	0.0	26.98	SmallLeak	10	Line_1	Location_1

(b)

2) Proses learning

Proses pembelajaran memiliki peran untuk dapat memberikan keputusan kondisi pipa mengalami kebocoran atau tidak. Gambaran tahapan proses pembelajaran dapat dilihat pada gambar 4.



Gambar 4. Proses pembelajaran untuk klasifikasi ukuran dan lokasi kebocoran

Raw dataset dan process dataset dilakukan pembelajaran oleh algoritma machine learning. Tahapan selanjutnya data dilakukan pre-processing, dan dilanjutkan dengan metode

normalisasi data. Pada proses pembelajaran ini penelitian menggunakan teknik *Hold-Out Validation*, dimana dataset dibagi menjadi 75% *data training* dan 25 % *data testing*.

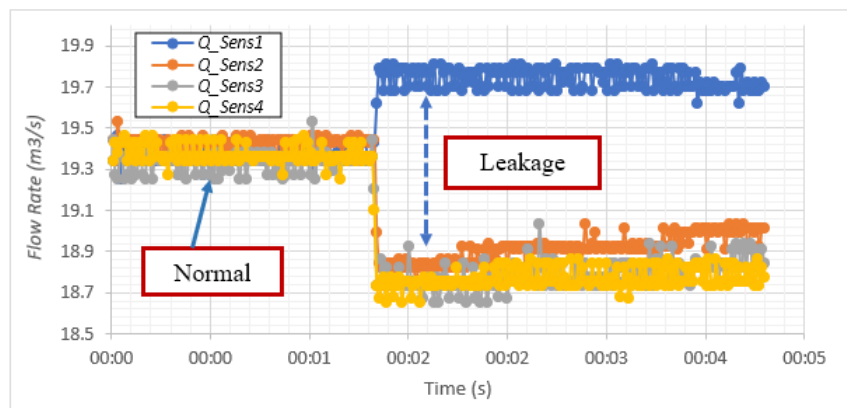
Pada penelitian, pelatihan model *machine learning* dilakukan pada beberapa model diantaranya adalah *Naïve Bayes*, *Multi Layer Perceptron (MLP)*, dan pengoptimalan penggunaan fungsi kernel algoritma SVM. Pelatihan pada model *machine learning* tersebut, menggunakan *library Scikit-Learn* [18]. Parameter-parameter yang digunakan dalam proses pembelajaran dibuat secara *default* yang ada di *library sklearn*, dimana ada pengecualian pada algoritma MLP Classifier dilakukan *setting parameter hidden layer* 10 dan maksimum iterasi sebanyak 1000 iterasi.

III. HASIL DAN PEMBAHASAN

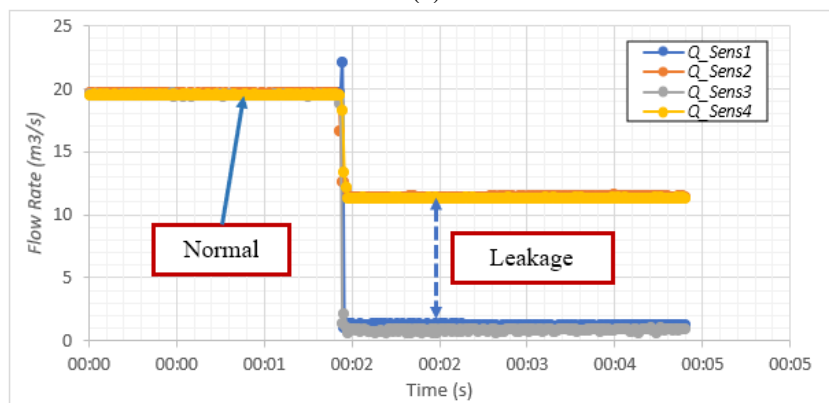
Eksperimen dalam penelitian ini dilakukan melalui beberapa bagian yaitu menganalisis data kebocoran pipa dari simulasi pipa, melakukan *pre-processing* pada kedua *dataset*, melakukan pembelajaran terhadap tiga model algoritma, dan menguji model apakah dapat mendeteksi kebocoran pipa dalam simulasi pipa.

A. Analisis Data

Pengujian awal dilakukan untuk menganalisa kebocoran pada posisi diantara node sensor 1 dan 2. Pada 5. (a). terlihat ada ketidakseimbangan antara $Q_Sens1 \neq Q_Sens2 = Q_Sens3 = Q_Sens4$ yang menandakan pipa ada kebocoran di lokasi 1. Sedangkan pada 5. (b). ketidaksetimbangan terjadi antara $Q_Sens1 = Q_Sens3 \neq Q_Sens2 = Q_Sens4$.



(a)



(b)

Gambar 5. Raw Data (a) Buka valve 20° (*Small Leak* pada *Location 1*) (b) Buka valve 60° (*Big Leak* pada *Location 1*)

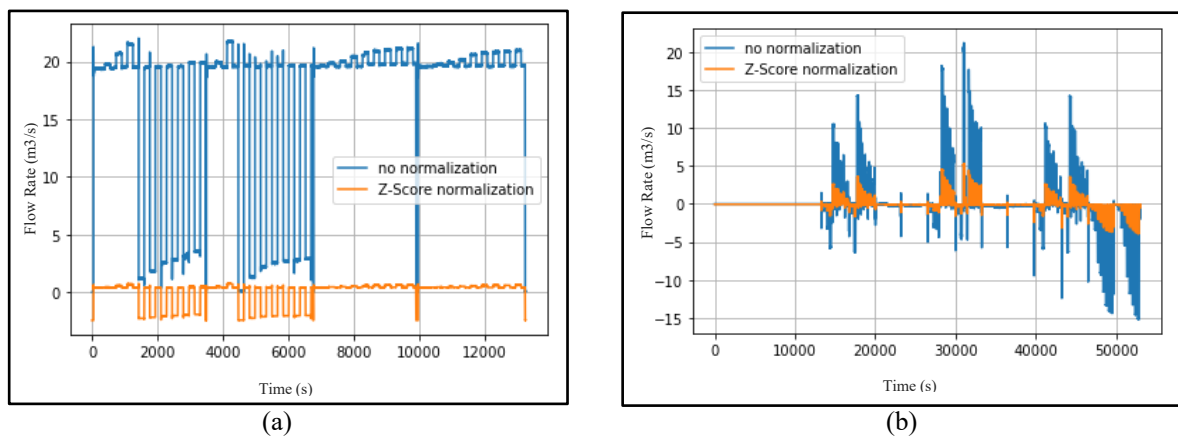
Data yang telah dianalisis kemudian diolah menjadi dua *dataset* yaitu *raw dataset* dan *process dataset*. Tahapan selanjutnya data dipersiapkan dengan melakukan beberapa teknik, yaitu menangani data yang hilang dan menangani data kategori.

B. Metode Normalisasi Data

Parameter yang dianalisa pada penelitian ini adalah dua model *dataset* yang disusun yaitu *raw dataset* dan *process dataset*. Metode normalisasi data adalah suatu metode untuk membuat suatu dataset memiliki skala yang sama, sehingga tidak ada variabel data yang mendominasi variabel data lainnya. Normalisasi *Z-score* adalah teknik normalisasi yang umum digunakan untuk penskalaan fitur. Teknik ini mengubah semua *data* pada skala standar dengan rata-rata nol dan standar deviasi satu. Persamaan normalisasi skor *z* dapat ditulis *Z-score normalization* pada persamaan nomor 5.

$$Z = \frac{x - \mu}{s} \tag{5}$$

Dimana, *Z* adalah data dari hasil *Z-score normalization*, *x* adalah data asli dari *dataset*, μ adalah rata-rata dari data, dan *s* merupakan standar deviasi dari data-data tersebut. Hasil dari normalisasi pada kedua *dataset* pada penelitian ini dapat dilihat pada gambar 6.



Gambar 6. Hasil *Z-score Normalization* (a). *Raw Dataset*, (b). *Process Dataset*

C. Eksperimen Model menggunakan Hold-Out Validation:

Untuk melihat apakah model algoritma SVM cocok dalam mengklasifikasikan ukuran dan lokasi kebocoran pipa. Akurasi merupakan ratio dari prediksi yang benar (positif dan negatif) dengan keseluruhan data dirumuskan dalam persamaan nomor 6.

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \tag{6}$$

True Positives (TP) mewakili kelas situasi aktual yang dapat diprediksi dengan benar dan akurat. *True Negatif* (TN) merepresentasikan data negatif yang diprediksi benar. *False Positives* (FP) mewakili kelas situasi aktual yang tidak dapat diprediksi dengan benar. *False Negative* (FN) mewakili kelas yang benar yang prediksinya salah.

Recall dapat dijelaskan sebagai rasio jumlah total contoh positif yang diklasifikasikan dengan benar dibagi dengan jumlah total sampel positif. Nilai *recall* yang tinggi menunjukkan kelas diprediksi secara akurat. Peramaan nomor 7 menunjukkan nilai *recall*:

$$Recall = \frac{TP}{TP+FN} \tag{7}$$

Sedangkan untuk Presisi dapat dihitung dengan menggunakan persamaan nomor 8,

$$presisi = \frac{TP}{TP+FP} \tag{8}$$

Dan *F-measure* dapat dihitung dengan menggunakan persamaan nomor 9,

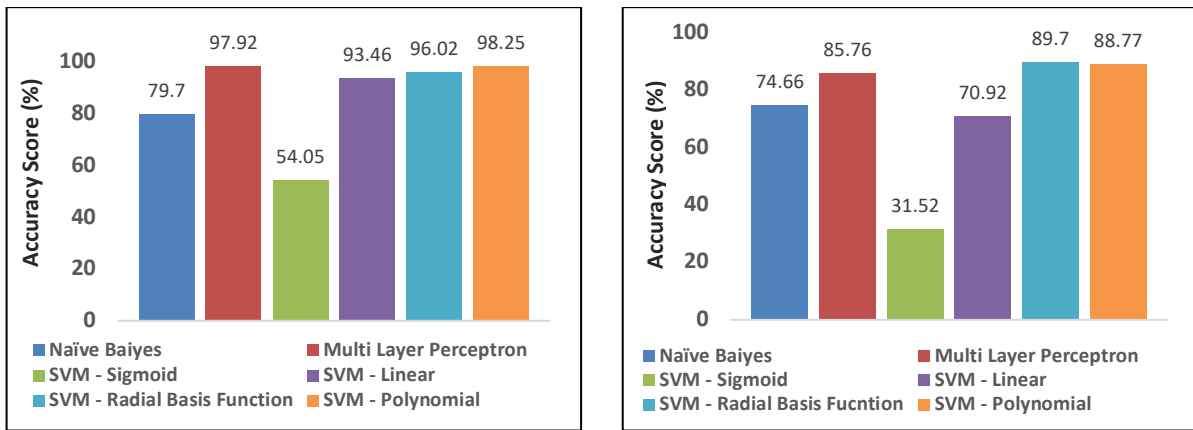
$$F - measure = 2 x \frac{Presisi \times Recall}{Presisi+Recall} \tag{9}$$

D. Eksperimen Sistem Deteksi Kebocoran:

Penelitian ini menghasilkan dua model klasifikasi yaitu memprediksi ukuran dan lokasi kebocoran pipa. Model tersebut diujikan pada kedua model tipe dataset yaitu *raw dataset* dan *process dataset*. Kedua model klasifikasi melalui beberapa tahapan proses dan dilakukan perbandingan kepada beberapa model algoritma seperti *Naïve Bayes*, *Multi Layer Perceptron* (MLP), dan *SVM* dengan empat fungsi kernel.

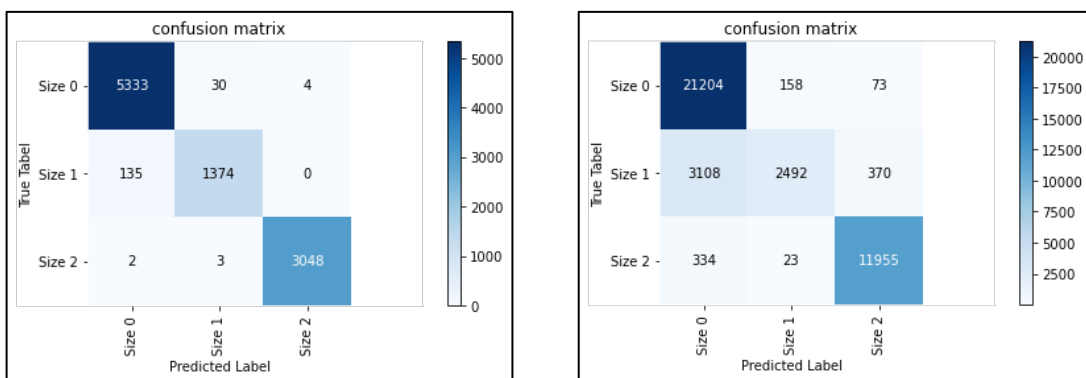
1. Prediksi Ukuran Kebocoran

Penelitian ini menghitung akurasi, *recall*, presisi, dan *F-measure* dari suatu model algoritma untuk memprediksi skenario ukuran kebocoran. Gambar 7 menunjukkan perbandingan nilai akurasi dari beberapa model algoritma dan optimasi penggunaan fungsi kernel dalam mengklasifikasikan ukuran kebocoran menggunakan *raw dataset* dan *process dataset*. Pada pembelajaran ukuran kebocoran fungsi kernel dengan akurasi tertinggi terlihat pada *raw dataset* menggunakan fungsi kernel *polynomial* dengan performa nilai akurasi sebesar 98,25%. Sedangkan pada *process dataset* yaitu sebesar 89,7%.



Gambar 7. Skor Akurasi Prediksi Ukuran Kebocoran, (a) pada *Raw dataset*, (b) pada *Process dataset*

Untuk melihat hasil prediksi tertinggi baik *raw dataset* dan *process dataset* terlihat pada *confusion matrix* pada gambar 8. Dari hasil yang didapat untuk kedua tipe dataset penggunaan model normalisasi data dan fitur kernel dapat meningkatkan performa akurasi model algoritma.



Gambar 8. *Confusion matrix* hasil klasifikasi, (a) *raw data* kernel polynomial, (b) *process data* kernel RBF

Selain parameter akurasi, parameter lain seperti *recall*, presisi dan *F-measure* ditunjukkan pula performanya untuk masing – masing klasifikasi ukuran kebocoran pada

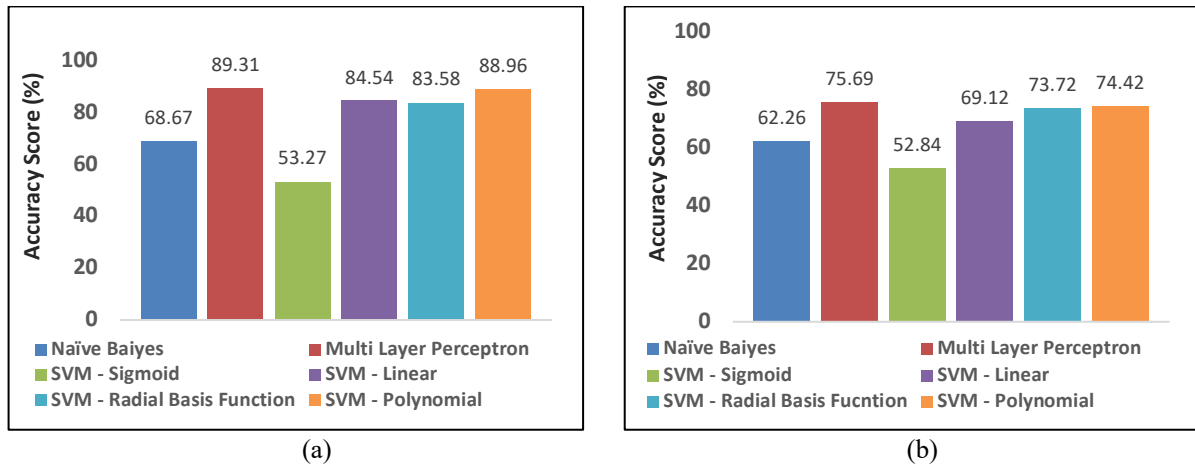
Tabel V. Pada *raw dataset*, hasil yang diperoleh untuk *recall* sebesar 99,1%, *presisi* sebesar 99,8% dan *F-measure* 99,5%. Hasil pada *process dataset*, *recall* sebesar 94,4%, *presisi* sebesar 95,4% dan *F-measure* 94,6%.

TABEL V
RECALL, PRESISI DAN F-MEASURE UNTUK UKURAN KEBOCORAN PIPA

Parameter		No Leak (0)	Small Leak (1)	Big Leak (2)	Rata-rata
Raw Dataset	Recall	0.975	0.998	1.000	0.991
	Presisi	0.994	1.000	0.999	0.998
	F-Measure	0.985	0.999	1.000	0.995
Process Dataset	Recall	0.872	0.991	0.970	0.944
	Presisi	0.993	0.871	0.998	0.954
	F-Measure	0.928	0.927	0.984	0.946

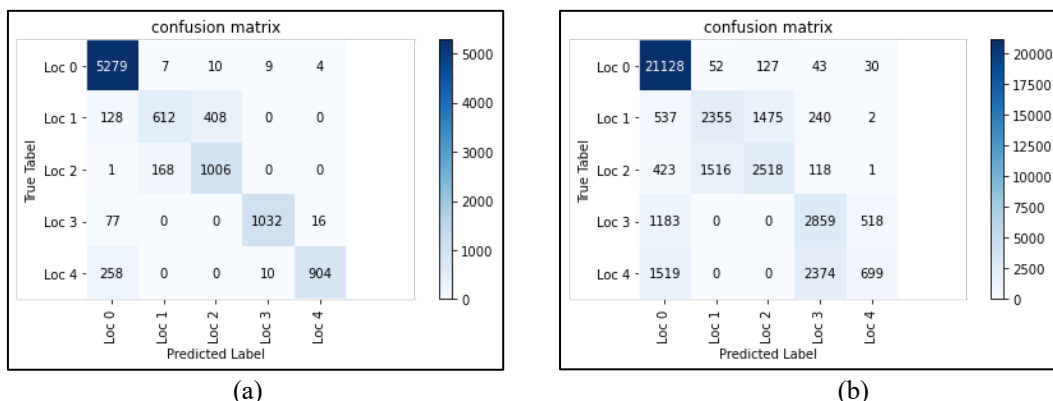
2. *Prediksi lokasi kebocoran*

Pada bagian pengklasifikasian lokasi kebocoran, skenario penelitian memperlakukan langkah metode pembelajaran yang sama seperti mendeteksi ukuran kebocoran. Gambar 9 menunjukkan nilai akurasi penggunaan model normalisasi data *z-score* dan optimalisasi penggunaan fungsi kernel pada algoritma SVM.



Gambar 9. Skor Akurasi Prediksi Lokasi Kebocoran, (a) *Raw dataset*, (b) *Process dataset*

Proses pembelajaran lokasi kebocoran terlihat pada *raw dataset* menunjukkan akurasi tertinggi pada MLP sebesar 89,31% sedangkan pada SVM kernel *polynomial* dengan performa akurasi sebesar 88,96% dengan selisih sebesar 0,35%, pada *process dataset* sebesar MP 75,69% dan SVM kernel *polynomial* sebesar 74,42% selisih cukup besar 1,27%. Hasil prediksi dari hasil pembelajaran menggunakan fungsi kernel SVM, prediksi lokasi kebocoran dapat terlihat pada *confusion matrix* pada gambar 10.



Gambar 10. Confusion matrix hasil klasifikasi fungsi kernel polynomial, (a) *raw dataset*, (b) *process dataset*

Terlihat pula hasil *recall*, presisi dan *F-measure* untuk lokasi kebocoran pipa pada Tabel VI. Pada *raw dataset* hasil yang diperoleh untuk *recall* sebesar 94,7%, *presisi* sebesar 91,5% dan *F-measure* 92,8%. Hasil pada *process dataset* *recall* sebesar 74,1%, *presisi* sebesar 72,8% dan *F-measure* 71,3%.

TABEL VI
RECALL, PRESISI DAN F-MEASURE UNTUK LOKASI KEBOCORAN PIPA

Parameter		No_Leak (0)	Location 1 (1)	Location 2 (2)	Location 3 (3)	Location 4 (4)	Rata-rata
Raw Dataset	Recall	0.976	0.785	1.000	0.990	0.983	0.947
	Presisi	0.999	0.600	1.000	0.985	0.989	0.915
	F-measure	0.987	0.680	1.000	0.988	0.986	0.928
Process Dataset	Recall	0.975	0.608	1.000	0.546	0.574	0.741
	Presisi	0.998	0.615	0.955	0.847	0.227	0.728
	F-measure	0.986	0.612	0.977	0.664	0.326	0.713

IV. KESIMPULAN

Penelitian ini bertujuan untuk mendapatkan model dataset, metode normalisasi data, dan model algoritma *machine learning* yang dapat mendeteksi ukuran dan lokasi kebocoran pada pipa. Dataset yang diuji dua tipe data yaitu *raw dataset* dan *process dataset* kemudian dilakukan normalisasi data menggunakan metode z-score. Model algoritma yang dilatih menggunakan *Naïve Bayes*, *Multi Layer Perceptron* (MLP), dan SVM dengan empat fungsi kernel seperti *sigmoid*, *radial basis function*, *linear*, dan *polynomial*. Masing – masing model algoritma dilatih sehingga dapat melakukan klasifikasi ukuran kedalam tiga klasifikasi yaitu *no leak*, *small leak* dan *big leak*, sedangkan untuk lokasi kebocoran kedalam lima klasifikasi *no leak*, *location 1*, *location 2*, *location 3* dan *location 4*. Pada klasifikasi ukuran kebocoran terlihat bahwa tingkat akurasi yang tertinggi diperoleh pada *raw dataset* daripada *process dataset*. Sedangkan pada model algoritma antara *naïve bayes*, *MLP* dan *SVM*, akurasi tertinggi sebesar 98,25%, *recall* 99,1%, presisi 99,8%, dan *F-measure* 99,5%. menggunakan *SVM* dengan fungsi kernel *polynomial*. Pada klasifikasi lokasi kebocoran terlihat pula *raw dataset* memiliki akurasi yang lebih baik dibandingkan *process dataset*. Akurasi tertinggi diperoleh *MLP* sebesar 89,31%, untuk *SVM* kernel *polynomial* nilai akurasi sebesar 88,96%, *recall* 94,7%, presisi 91,5% dan *F-measure* 92,8%, sehingga selesih terhadap *MLP* sebesar 0,35%. Sedangkan pada *SVM* kernel *sigmoid* dan *naïve bayes* tertinggal cukup jauh. Model yang terbaik dari pembelajaran kemudian ditanamkan pada *gateway* sehingga sistem IoT dapat mendeteksi ukuran dan lokasi kebocoran pipa dan menginformasikan pengguna secara *realtime* melalui website dan aplikasi komunikasi seperti telegram.

REFERENSI

- [1] S. K. Mandal, F. T. S. Chan, and M. K. Tiwari, "Leak detection of pipeline: An integrated approach of rough set theory and artificial bee colony trained SVM," *Expert Systems with Applications*, vol. 39, no. 3, pp. 3071–3080, 2012, doi: 10.1016/j.eswa.2011.08.170.
- [2] J. Mashford, D. de Silva, D. Marney, and S. Burn, "An approach to leak detection in pipe networks using analysis of monitored pressure values by support vector machine," *NSS 2009 - Network and System Security*, pp. 534–539, 2009, doi: 10.1109/NSS.2009.38.
- [3] N. Mashhadi, I. Shahrour, N. Attoue, J. el Khattabi, and A. Aljer, "Use of machine learning for leak detection and localization in water distribution systems," *Smart Cities*, vol. 4, no. 4, pp. 1293–1315, Dec. 2021, doi: 10.3390/smartcities4040069.
- [4] H. W. Li, J. L. Yan, W. Wu, and L. Mou, "Research on water leakage prediction of urban water supply network based on grey GM (0, N) model," *2011 International Conference*

- on Networking, Sensing and Control, ICNSC 2011*, no. April, pp. 388–392, 2011, doi: 10.1109/ICNSC.2011.5874884.
- [5] M. S. Islam, R. Sadiq, M. J. Rodriguez, A. Francisque, H. Najjaran, and M. Hoorfar, “Leakage detection and location in water distribution systems using a fuzzy-based methodology,” *Urban Water Journal*, vol. 8, no. 6, pp. 351–365, 2011, doi: 10.1080/1573062X.2011.617829.
- [6] A. Rojik, Endroyono, and A. N. Irfansyah, “Water Pipe Leak Detection using the k-Nearest Neighbor Method,” *Proceedings - 2019 International Seminar on Intelligent Technology and Its Application, ISITIA 2019*, pp. 393–398, 2019, doi: 10.1109/ISITIA.2019.8937190.
- [7] Y. Liu, X. Ma, Y. Li, Y. Tie, Y. Zhang, and J. Gao, “Water pipeline leakage detection based on machine learning and wireless sensor networks,” *Sensors (Switzerland)*, vol. 19, no. 23, pp. 1–21, 2019, doi: 10.3390/s19235086.
- [8] J. C. van der Walt, P. S. Heyns, and D. N. Wilke, “Pipe network leak detection: comparison between statistical and machine learning techniques,” *Urban Water Journal*, vol. 15, no. 10, pp. 953–960, 2018, doi: 10.1080/1573062X.2019.1597375.
- [9] R. H. Y. Perdana, H. Hudiono, and A. F. N. Luqmani, “Water Leak Detection and Shut-Off System on Water Distribution Pipe Network Using Wireless Sensor Network,” in *2019 International Conference on Advanced Mechatronics, Intelligent Manufacture and Industrial Automation (ICAMIMIA)*, Oct. 2019, pp. 297–301. doi: 10.1109/ICAMIMIA47173.2019.9223386.
- [10] Y. Baiji and P. Sundaravadivel, “ILoLeak-detect: An IoT-based LoRAWAN-enabled oil leak detection system for smart cities,” in *Proceedings - 2019 IEEE International Symposium on Smart Electronic Systems, iSES 2019*, Dec. 2019, pp. 262–267. doi: 10.1109/iSES47678.2019.00065.
- [11] J. Alves Coelho, A. Glória, and P. Sebastião, “Precise Water Leak Detection Using Machine Learning and Real-Time Sensor Data,” *IoT*, vol. 1, no. 2, pp. 474–493, Dec. 2020, doi: 10.3390/iot1020026.
- [12] J. A. Gómez-Camperos, E. E. Espinel-Blanco, and F. J. Regino-Ubarnes, “Diagnosis of horizontal pipe leaks using neural networks,” *Journal of Physics: Conference Series*, vol. 1388, no. 1, 2019, doi: 10.1088/1742-6596/1388/1/012032.
- [13] M. Henrie, P. Carpenter, and R. E. Nicholas, *Real-Time Transient Model-Based Leak Detection*. 2016. doi: 10.1016/b978-0-12-802240-5.00004-2.
- [14] G. Geiger, “State-of-the-art in leak detection and localization,” *Erdoel Erdgas Kohle*, vol. 122, no. 12, pp. 193–198, 2006.
- [15] Z. X. Yang and J. H. Zhong, “A hybrid EEMD-based SampEn and SVD for acoustic signal processing and fault diagnosis,” *Entropy*, vol. 18, no. 4, Apr. 2016, doi: 10.3390/e18040112.
- [16] S. Bakheet, “An SVM framework for malignant melanoma detection based on optimized HOG features,” *Computation*, vol. 5, no. 1, Mar. 2017, doi: 10.3390/computation5010004.
- [17] J. A. Coelho and A. Glória, “IoT Learning and Real-Time Sensor Data,” pp. 474–493, 2020, doi: 10.3390/iot1020026.
- [18] Scikit-Learn, “scikit-learn Machine Learning in Python,” 2022.