

Komparasi Algoritma *Random Forest* dan *SVM* dalam Klasifikasi Kondisi Aliran Air Berdasarkan Getaran MPU6050

Aknes Tasia Pratama, Ahmad Taqwa*, Lindawati

Sarjana Terapan Teknik Telekomunikasi, Jurusan Teknik Elektro, Politeknik Negeri Sriwijaya, Palembang, Indonesia.

Email: ¹aknestasia2279@gmail.com, ^{2*}taqwa@polsri.ac.id, ³lindawati@polsri.ac.id

Email Penulis Korespondensi: taqwa@polsri.ac.id*

Submitted: 06/06/2026; Accepted: 16/06/2026; Published: 30/06/2026

Abstrak– Ketersediaan informasi mengenai kondisi aliran air pada sistem distribusi masih menjadi permasalahan di beberapa wilayah, terutama pada jaringan yang tidak beroperasi secara kontinu. Penelitian ini bertujuan mengembangkan model klasifikasi kondisi aliran air pada pipa distribusi berdasarkan data getaran yang diperoleh menggunakan sensor MPU6050. Data getaran direpresentasikan melalui fitur akselerasi tiga sumbu (ax , ay , dan az) serta fitur statistik berupa *mean* dan *Root Mean Square* (RMS). *Dataset* yang digunakan terdiri atas 4.406 sampel yang dikelompokkan ke dalam dua kelas, yaitu kondisi air mengalir dan pipa kosong. Tahapan penelitian meliputi pra-pemrosesan data menggunakan *StandardScaler*, penanganan ketidakseimbangan kelas menggunakan *Synthetic Minority Over-sampling Technique* (SMOTE), serta pelatihan model menggunakan algoritma *Random Forest* dan *Support Vector Machine* (SVM). Hasil pengujian menunjukkan bahwa *Random Forest* memperoleh akurasi 95,92%, presisi 96,60%, *recall* 96,22%, dan *F1-score* 96,41%, sedangkan *Support Vector Machine* (SVM) memperoleh akurasi 95,80%, presisi 95,50%, *recall* 97,21%, dan *F1-score* 96,35%. Hasil tersebut menunjukkan bahwa kedua algoritma mampu mengklasifikasikan kondisi aliran air dengan baik, dengan *Random Forest* memberikan performa keseluruhan yang sedikit lebih unggul dibandingkan *Support Vector Machine* (SVM). Penelitian ini menunjukkan bahwa data getaran dari sensor MPU6050 berpotensi digunakan sebagai solusi pemantauan kondisi aliran air secara *non-intrusif* pada sistem distribusi air.

Kata Kunci: MPU6050; *Machine Learning*; *Random Forest*; *Support Vector Machine*; Distribusi Air.

Abstract– The availability of information regarding water flow conditions in distribution systems remains a challenge in several areas, particularly in networks that do not operate continuously. This study aims to develop a classification model for water flow conditions in distribution pipes based on vibration data obtained using an MPU6050 sensor. The vibration data are represented by three-axis acceleration features (ax , ay , and az) and statistical features, namely mean and Root Mean Square (RMS). The dataset consists of 4,406 samples categorized into two classes: flowing water and empty pipe conditions. The research stages include data preprocessing using *StandardScaler*, handling class imbalance using the *Synthetic Minority Over-sampling Technique* (SMOTE), and model training using *Random forest* and *Support Vector Machine* (SVM) algorithms. The experimental results show that *Random forest* achieved an accuracy of 95.92%, precision of 96.60%, recall of 96.22%, and F1-score of 96.41%, while *Support Vector Machine* (SVM) achieved an accuracy of 95.80%, precision of 95.50%, recall of 97.21%, and F1-score of 96.35%. These results indicate that both algorithms can effectively classify water flow conditions, with *Random forest* providing slightly better overall performance compared to *Support Vector Machine* (SVM). This study demonstrates that vibration data from the MPU6050 sensor has the potential to be used as a non-intrusive solution for monitoring water flow conditions in water distribution systems.

Keywords: MPU6050; *Machine Learning*; *Random Forest*; *Support Vector Machine*; Water Distribution.

1. PENDAHULUAN

Ketersediaan air bersih menjadi salah satu faktor fundamental yang memengaruhi kualitas hidup masyarakat karena berkaitan langsung dengan aspek kesehatan, produktivitas, dan keberlangsungan aktivitas sehari-hari. Ketersediaan dan distribusi air bersih yang memadai juga menjadi salah satu indikator utama dalam pencapaian *Sustainable Development Goals* (SDGs) [1]. Di Indonesia, persentase rumah tangga yang memiliki akses terhadap sumber air minum layak terus mengalami peningkatan dari tahun ke tahun berdasarkan Statistik Lingkungan Hidup Indonesia 2024 [2]. Pada beberapa wilayah perdesaan, distribusi air masih dilakukan secara bergilir dengan jadwal yang tidak menentu sehingga masyarakat kesulitan mengetahui waktu kedatangan air pada saluran distribusi. Ketersediaan dan distribusi air yang andal memiliki peran penting dalam mendukung aktivitas sosial ekonomi masyarakat serta pembangunan wilayah secara berkelanjutan, terutama pada daerah yang masih menghadapi keterbatasan infrastruktur air bersih [3]. Oleh karena itu, dibutuhkan suatu pendekatan yang mampu memberikan informasi mengenai kondisi aliran air secara lebih efektif tanpa memerlukan pengamatan langsung secara terus-menerus.

Penelitian ini mengambil studi kasus pada sistem distribusi air di Desa Lubuk Raman. Lokasi tersebut dipilih karena masyarakat masih menghadapi ketidakpastian waktu distribusi air sehingga diperlukan metode yang dapat membantu mengenali keberadaan aliran air secara otomatis. Pada praktiknya, kondisi aliran air umumnya diketahui melalui pemeriksaan langsung pada keran atau saluran distribusi. Metode tersebut relatif sederhana, namun kurang efisien apabila dilakukan secara berulang dan tidak mampu memberikan informasi secara berkelanjutan. Oleh sebab itu, diperlukan metode yang mampu mengidentifikasi kondisi aliran air berdasarkan karakteristik fisik tertentu yang muncul selama proses distribusi berlangsung.

Salah satu karakteristik yang dapat dimanfaatkan adalah pola getaran pada permukaan pipa distribusi. Ketika air mengalir di dalam pipa, interaksi antara fluida dan dinding pipa menghasilkan pola getaran yang berbeda dibandingkan kondisi pipa kosong [4]. Pendekatan berbasis getaran memiliki keunggulan karena bersifat *non-intrusif* sehingga sensor dapat dipasang pada bagian luar pipa tanpa mengubah struktur sistem distribusi maupun meningkatkan risiko kebocoran [5]. Dengan memanfaatkan data getaran sebagai sumber informasi, proses identifikasi kondisi aliran dapat dilakukan tanpa melakukan perubahan yang signifikan pada infrastruktur distribusi air yang telah tersedia.

Tantangan utama dalam pemanfaatan data getaran untuk identifikasi kondisi aliran air terletak pada kompleksitas pola sinyal yang dihasilkan. Getaran yang berasal dari kendaraan, aktivitas manusia, maupun gangguan mekanis lainnya dapat menghasilkan pola yang menyerupai karakteristik getaran akibat aliran air. Kondisi tersebut menyebabkan pendekatan berbasis ambang batas tetap (*fixed threshold*) menjadi kurang efektif karena sulit mengakomodasi variasi kondisi yang terjadi di lapangan [6]. Sebab itu, dibutuhkan cara analisis yang bisa mengenali pola dengan lebih fleksibel dan tidak hanya mengandalkan aturan-aturan sederhana yang sudah ditetapkan sejak awal. Pemanfaatan teknologi pintar (*smart technologies*) dalam pengelolaan sumber daya air juga semakin berkembang karena mampu menyediakan informasi yang lebih akurat dan mendukung proses pengambilan keputusan yang lebih efektif [7].

Kemajuan metode komputasi berbasis data memungkinkan informasi yang berasal dari sensor diproses secara lebih adaptif. Melalui proses pembelajaran dari data historis, sistem dapat mengenali karakteristik tertentu dan menghasilkan keputusan klasifikasi secara otomatis tanpa memerlukan aturan berbasis ambang yang kaku. Melalui proses pelatihan menggunakan data historis, algoritma *machine learning* mampu mempelajari karakteristik masing-masing kelas dan melakukan klasifikasi secara otomatis berdasarkan pola yang terbentuk [8], [4]. Selain itu, integrasi teknologi *Internet of Things (IoT)* dan *machine learning* memungkinkan sistem pemantauan distribusi air dilakukan secara otomatis, berkelanjutan, dan *real-time* [9], [10]. Dalam penelitian ini digunakan dua algoritma klasifikasi yang banyak diterapkan pada permasalahan pengenalan pola, yaitu *Random Forest* dan *Support Vector Machine (SVM)*. Salah satu pendekatan yang banyak digunakan untuk permasalahan klasifikasi adalah *Random Forest*, yaitu metode yang menggabungkan hasil prediksi dari sejumlah model pohon keputusan. Pendekatan kolektif tersebut membantu meningkatkan konsistensi prediksi sekaligus mengurangi pengaruh data yang bersifat menyimpang. Berbeda dengan *Random Forest*, *SVM* membentuk batas keputusan yang dirancang untuk memberikan pemisahan terbaik antar kelompok data sehingga objek dari kelas yang berbeda dapat dibedakan secara lebih jelas [11], [12].

Berbagai penelitian sebelumnya telah menunjukkan keberhasilan penerapan *machine learning* pada sistem distribusi air. Liu et al. [4] melakukan kajian sistematis terhadap model *machine learning* berbasis teknologi *vibroakustik* untuk deteksi kebocoran air dan menyimpulkan bahwa data getaran memiliki potensi besar sebagai sumber informasi kondisi pipa. Rajgopal et al. [13] mengembangkan sistem deteksi kebocoran berbasis *Wireless Sensor Network (WSN)* untuk meningkatkan pengelolaan distribusi air. Hasil penelitian menunjukkan bahwa kombinasi sensor dan algoritma klasifikasi mampu meningkatkan kemampuan pemantauan jaringan distribusi. Pourmehrani et al. [6] mengembangkan perangkat deteksi aliran air berbasis analisis suara dan *machine learning* yang mampu membedakan kondisi aliran dan kebocoran pada pipa distribusi. Selain itu, Widyatma et al. [5] mengimplementasikan sistem deteksi kebocoran pipa berbasis *IoT* yang menunjukkan bahwa teknologi digital dapat meningkatkan efektivitas pemantauan jaringan perpipaan. Pagano et al. [14] juga menjelaskan bahwa penerapan *IoT* skala besar pada sistem distribusi air memiliki potensi besar untuk mendukung pengelolaan jaringan perpipaan secara luas dan berkelanjutan.

Meskipun berbagai penelitian tersebut telah menunjukkan keberhasilan penggunaan *machine learning* pada sistem distribusi air, sebagian besar penelitian masih berfokus pada deteksi kebocoran menggunakan sensor akustik, sensor intrusif, atau jaringan sensor khusus. Penelitian yang memanfaatkan data getaran *non-intrusif* dari sensor MPU6050 untuk mengklasifikasikan kondisi aliran air masih relatif terbatas. Selain itu, sebagian besar penelitian hanya berfokus pada deteksi kebocoran atau identifikasi kondisi tertentu tanpa melakukan analisis komparatif terhadap algoritma klasifikasi yang digunakan. Hingga saat ini masih sedikit penelitian yang membandingkan performa algoritma *Random Forest* dan *SVM* dalam mengklasifikasikan kondisi aliran air berdasarkan data getaran pipa distribusi. Kondisi tersebut menunjukkan adanya *research gap* yang menjadi dasar pelaksanaan penelitian ini.

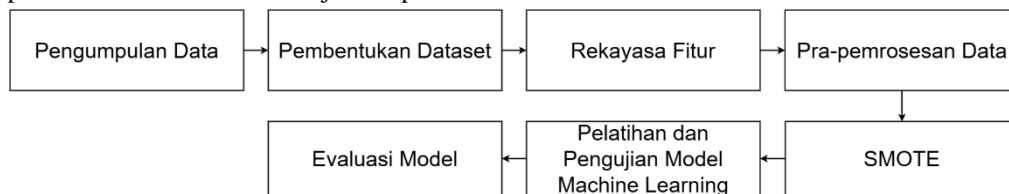
Berdasarkan *research gap* tersebut, pemanfaatan data getaran untuk identifikasi aliran air belum banyak dieksplorasi secara spesifik pada sistem distribusi air masyarakat perdesaan. Oleh karena itu, tujuan utama dari penelitian ini adalah membandingkan dan mengevaluasi kinerja algoritma klasifikasi *Random Forest* dan *Support Vector Machine (SVM)* dalam mengklasifikasikan kondisi aliran air berdasarkan data getaran sensor MPU6050. Melalui analisis komparatif ini, penelitian ditujukan untuk menemukan model dengan tingkat akurasi tertinggi dalam membedakan kondisi pipa beraliran dan pipa kosong. Hasil akhir dari penelitian ini diharapkan dapat menjadi landasan fundamental dalam pengembangan sistem peringatan dini distribusi air yang otomatis, *non-intrusif*, dan responsif.

2. METODOLOGI PENELITIAN

Alur penelitian dirancang dalam beberapa tahap yang saling berurutan, dimulai dari akuisisi sinyal getaran hingga evaluasi model klasifikasi yang dibangun menggunakan pendekatan *machine learning*. Fokus utama penelitian adalah membandingkan kemampuan algoritma *Random Forest* dan *Support Vector Machine (SVM)* dalam mengklasifikasikan kondisi aliran air berdasarkan data getaran yang didapat dari pipa.

2.1 Tahapan Penelitian

Penelitian ini berfokus pada analisis performa algoritma klasifikasi dalam mengenali pola getaran pada pipa distribusi air menggunakan pendekatan *Machine learning*. Tahapan penelitian dirancang dalam bentuk *machine learning pipeline* yang terdiri atas pengumpulan data, pembentukan *dataset*, rekayasa fitur (*feature engineering*), pra-pemrosesan data, penanganan ketidakseimbangan kelas, pelatihan model, dan evaluasi performa model. Pendekatan *pipeline* dipilih karena mampu menghasilkan proses pengolahan data yang sistematis, terstruktur, serta meminimalkan kemungkinan terjadinya *data leakage* selama proses pelatihan dan pengujian model [15]. Tahapan penelitian secara umum ditunjukkan pada Gambar 1.



Gambar 1. Diagram Alir Tahapan Penelitian

Alur kerja penelitian pada Gambar 1 menggambarkan rangkaian proses yang dilakukan mulai dari proses pengambilan data getaran melalui sensor MPU6050 yang dipasang di luar pipa distribusi air. Data yang diperoleh kemudian disimpan dalam bentuk *dataset* dan diberi label berdasarkan kondisi aliran air. *Dataset* selanjutnya diproses melalui tahapan *feature engineering* untuk menghasilkan fitur yang merepresentasikan karakteristik getaran pipa. Selanjutnya dilakukan pra-pemrosesan data yang meliputi pembersihan data, pembagian *dataset*, normalisasi, dan penanganan ketidakseimbangan kelas. Data hasil pra-pemrosesan digunakan untuk melatih model *Random Forest* dan *SVM*. Tahap terakhir adalah evaluasi performa model menggunakan metrik klasifikasi untuk menentukan algoritma yang memiliki performa terbaik.

2.2 Pengumpulan Data

Penelitian ini menggunakan *dataset* yang berasal dari hasil pengukuran kondisi aliran air pada pipa distribusi. *Dataset* memuat data percepatan pada tiga sumbu pengukuran (*ax*, *ay*, dan *az*) beserta label kondisi aliran yang digunakan sebagai target klasifikasi. Data yang diperoleh selanjutnya digunakan sebagai masukan pada tahapan rekayasa fitur, pra-pemrosesan data, pelatihan model, dan evaluasi performa algoritma *machine learning*.

Pengambilan data dilakukan pada dua kondisi operasional yang berbeda, yaitu kondisi saat air mengalir dan kondisi saat pipa kosong. Setiap sampel yang diperoleh kemudian diberikan label sesuai dengan kondisi aktual pipa. Proses pelabelan data ditunjukkan pada Tabel 1.

Tabel 1. Kondisi Pengambilan Data

Label	Kondisi
1	Air Mengalir
0	Pipa Kosong

Proses pelabelan dilakukan secara langsung selama pengambilan data sehingga setiap sampel memiliki kelas target yang jelas untuk proses pelatihan model klasifikasi.

2.3 Pembentukan *Dataset*

Dataset yang digunakan dalam penelitian merupakan hasil pengukuran getaran pipa distribusi air menggunakan sensor MPU6050 yang telah melalui proses pembersihan data (*data cleaning*). Setelah penghapusan data yang tidak valid dan data yang memiliki nilai kosong (*missing value*), diperoleh sebanyak 4.406 sampel data yang siap digunakan untuk proses pelatihan dan pengujian model.

Dataset terdiri atas lima fitur masukan (*input features*) dan satu variabel target (*target variable*). Struktur *Dataset* yang digunakan ditunjukkan pada Tabel 2.

Tabel 2. Struktur *Dataset*

Fitur	Deskripsi
ax	Nilai Percepatan Sumbu X
ay	Nilai Percepatan Sumbu Y
az	Nilai Percepatan Sumbu Z
mean	Nilai Rata-Rata Getaran
rms	Nilai <i>Root Mean Square</i>
label	Status Kondisi Aliran Air

Variabel target pada penelitian ini adalah kolom label, sedangkan fitur yang digunakan sebagai masukan model terdiri atas nilai *ax*, *ay*, *az*, *mean*, dan *rms*. Kombinasi fitur tersebut digunakan untuk menggambarkan karakteristik getaran yang muncul pada pipa distribusi dalam berbagai kondisi aliran air. Distribusi jumlah sampel pada masing-masing kelas ditunjukkan pada Tabel 3.

Tabel 3. Distribusi Kelas *Dataset*

Label	Kondisi	Jumlah Sampel
1	Air Mengalir	2.509
0	Pipa Kosong	1.897
Total	-	4.406

Berdasarkan Tabel 3 dapat diketahui bahwa jumlah sampel pada kelas air mengalir lebih banyak dibandingkan kelas pipa kosong. Perbedaan jumlah sampel ini menunjukkan adanya ketidakseimbangan kelas (*class imbalance*) yang berpotensi menyebabkan model lebih dominan mempelajari karakteristik kelas mayoritas. Oleh karena itu, diperlukan metode penanganan ketidakseimbangan kelas pada tahap pra-pemrosesan data.

2.4 Rekayasa Fitur (*Feature engineering*)

Rekayasa fitur dilakukan untuk menghasilkan parameter yang mampu merepresentasikan karakteristik getaran pipa secara lebih baik dibandingkan data akselerasi mentah. Pada penelitian ini digunakan dua fitur statistik utama, yaitu *mean* dan *Root Mean Square* (RMS). Kedua fitur tersebut banyak digunakan pada analisis sinyal getaran karena mampu menggambarkan kecenderungan amplitudo dan energi getaran secara efektif [16].

2.4.1 Resultan Magnitudo

Langkah pertama dalam proses rekayasa fitur adalah menghitung nilai resultan magnitudo percepatan dari ketiga sumbu sensor MPU6050. Nilai ini digunakan untuk merepresentasikan total percepatan tanpa dipengaruhi orientasi pemasangan sensor. Resultan magnitudo dihitung menggunakan Persamaan (1).

$$M = \sqrt{(ax^2 + ay^2 + az^2)} \tag{1}$$

dengan:

M = resultan magnitudo percepatan

ax = percepatan sumbu X

ay = percepatan sumbu Y

az = percepatan sumbu Z

2.4.2 Perhitungan *Mean*

Nilai *mean* digunakan untuk menggambarkan rata-rata amplitudo getaran dalam satu jendela pengamatan (*window*). Nilai ini memberikan informasi mengenai kecenderungan umum intensitas getaran yang terjadi pada pipa distribusi. Perhitungan *mean* dilakukan menggunakan Persamaan (2).

$$mean = (1/N) \times \sum_{i=1}^N M_i \tag{2}$$

dengan:

N = jumlah sampel dalam satu *window*

M_i = nilai resultan magnitudo pada sampel ke-*i*

2.4.3 Perhitungan *Root Mean Square* (RMS)

Selain nilai rata-rata, penelitian ini menggunakan parameter RMS yang mampu menggambarkan energi efektif dari sinyal getaran. Nilai RMS memiliki sensitivitas yang lebih tinggi terhadap perubahan amplitudo sehingga sangat efektif digunakan untuk membedakan kondisi pipa kosong dan kondisi air mengalir [16]. Nilai RMS dihitung menggunakan Persamaan (3).

$$RMS = \sqrt{\left(\frac{1}{N}\right) \times \sum_{i=1}^N M_i^2} \tag{3}$$

dengan:

N = jumlah sampel dalam satu *window*

M_i = nilai resultan magnitudo pada sampel ke-*i*

2.5 Pra-pemrosesan Data dan *Machine learning Pipeline*

Kualitas data menjadi faktor penting dalam pembentukan model klasifikasi yang andal. Oleh karena itu, *dataset* yang telah dikumpulkan menjalani beberapa tahapan persiapan agar karakteristiknya lebih seragam dan sesuai untuk proses pelatihan algoritma. Tahapan ini sangat penting karena kualitas data yang digunakan sangat berpengaruh terhadap kemampuan model dalam mengenali pola dan melakukan klasifikasi dengan tepat. Dalam penelitian ini, tahap pra-pemrosesan data mencakup membersihkan data, membagi *dataset*, melakukan normalisasi pada fitur, serta mengatasi masalah ketidakseimbangan kelas. Urutan langkah-langkah diterapkan dengan ketat agar tidak terjadi kebocoran data yang bisa menyebabkan penilaian kinerja model menjadi terlalu optimis [17].

2.5.1 Pembersihan Data

Proses pembersihan data (*data cleaning*) dilakukan dengan menghapus atribut yang tidak digunakan dalam proses klasifikasi, seperti *timestamp* dan atribut identifikasi lainnya. Selain itu dilakukan pemeriksaan terhadap nilai kosong (*missing value*) menggunakan metode *dropna()* sehingga hanya data yang lengkap yang digunakan dalam proses pelatihan model.

Setelah proses pembersihan data selesai dilakukan, diperoleh sebanyak 4.406 sampel data yang siap digunakan untuk tahap selanjutnya

2.5.2 Pembagian *Dataset*

Dataset dibagi menjadi data latih (*training data*) dan data uji (*testing data*) menggunakan metode *train-test split* dengan rasio 80:20. Pembagian dilakukan menggunakan pendekatan *stratified sampling* dengan parameter *random_state = 42* sehingga proporsi kelas pada data latih dan data uji tetap terjaga. Hasil pembagian *dataset* ditunjukkan pada Tabel 4.

Tabel 4. Hasil Pembagian *Dataset*

<i>Dataset</i>	Jumlah Sampel
Data Latih	3.524
Data Uji	882
Total	4.406

Distribusi kelas setelah proses pembagian *dataset* ditunjukkan pada Tabel 5.

Tabel 5. Distribusi Kelas Setelah Pembagian *Dataset*

<i>Dataset</i>	Label 1	Label 0	Total
Data Latih	2.007	1.517	3.524
Data Uji	502	380	882

Tabel 5 menunjukkan bahwa distribusi kelas pada data latih dan data uji tetap proporsional terhadap distribusi *dataset* awal sehingga dapat merepresentasikan kondisi data yang sebenarnya.

2.5.3 Normalisasi Data

Setelah data dibagi, selanjutnya data dinormalisasi dengan metode *StandardScaler*. Normalisasi bertujuan agar semua fitur memiliki skala yang sama, sehingga setiap fitur memiliki rata-rata hampir nol dan standar deviasi hampir satu. Langkah ini menjadi tahapan krusial, terutama bagi algoritma *Support Vector Machine* (SVM), karena algoritma tersebut bekerja berdasarkan perhitungan jarak (*distance-based*) yang sangat sensitif terhadap perbedaan rentang nilai antar fitur [18], [19]. Normalisasi dilakukan menggunakan Persamaan (4).

$$z = \frac{x - \mu}{\sigma} \tag{4}$$

dengan:

z = nilai hasil normalisasi

x = nilai fitur

μ = rata-rata fitur

σ = standar deviasi fitur

2.6 Penanganan Ketidakseimbangan Kelas

Distribusi data latih menunjukkan ketidakseimbangan jumlah sampel antara kelas air mengalir dan pipa kosong. Kondisi ini dapat menyebabkan model lebih dominan mempelajari kelas mayoritas sehingga kemampuan mengenali kelas minoritas menjadi menurun [20]. Untuk mengatasinya, digunakan metode *Synthetic Minority Over-sampling Technique* (SMOTE), yaitu teknik yang menghasilkan sampel sintesis baru pada kelas minoritas berdasarkan informasi dari *nearest neighbors* sehingga distribusi data menjadi lebih seimbang [20]. Berbeda dengan *oversampling* biasa yang hanya menduplikasi data, SMOTE membentuk sampel baru melalui interpolasi sehingga dapat mengurangi risiko *overfitting*. Pada penelitian ini, SMOTE diterapkan setelah normalisasi dan hanya pada data latih untuk menjaga independensi data uji. Dampak penerapan SMOTE terhadap distribusi data dibahas pada bagian hasil dan pembahasan.

2.7 Pelatihan Model *Machine learning*

Tahap pemodelan dilakukan menggunakan dua pendekatan klasifikasi yang memiliki karakteristik berbeda. Algoritma pertama memanfaatkan kombinasi banyak pohon keputusan untuk menghasilkan prediksi kolektif, sedangkan algoritma kedua membangun bidang pemisah optimal pada ruang fitur. Kedua algoritma dipilih karena memiliki kemampuan yang baik dalam menangani permasalahan klasifikasi data sensor dan telah banyak digunakan pada penelitian berbasis pengenalan pola (*pattern recognition*) [11], [12].

2.7.1 *Random Forest*

Model *Random Forest* dibangun melalui kombinasi sejumlah pohon keputusan yang bekerja secara independen [21]. Prediksi akhir ditentukan berdasarkan hasil agregasi seluruh pohon sehingga keputusan yang dihasilkan cenderung lebih stabil terhadap variasi data. Pendekatan ini mampu meningkatkan stabilitas model dan mengurangi risiko *overfitting* dibandingkan penggunaan satu pohon keputusan tunggal [11].

Konfigurasi parameter *Random Forest* yang digunakan pada penelitian ini adalah $n_estimators = 100$, dan $random_state = 42$. Model *Random Forest* dilatih menggunakan data hasil normalisasi dan SMOTE, kemudian digunakan untuk melakukan prediksi pada data uji.

2.7.2 *Support Vector Machine*

Support Vector Machine (SVM) adalah algoritma klasifikasi yang bekerja dengan mencari garis pemisah terbaik (*hyperplane*) untuk membagi data dari berbagai kelas dengan jarak (*margin*) yang terbesar. Penelitian ini menggunakan fungsi kernel *Radial Basis Function* (RBF) karena kemampuannya dalam menangani hubungan tidak linier antar fitur yang sering muncul pada data getaran sensor [12]. Konfigurasi parameter SVM yang digunakan adalah Kernel RBF. Model SVM dilatih menggunakan data yang sama dengan *Random Forest* sehingga perbandingan performa kedua algoritma dapat dilakukan secara objektif pada kondisi eksperimen yang identik.

2.8 Evaluasi Kinerja Model

Kinerja model dinilai melalui analisis hubungan antara hasil prediksi dan kondisi aktual pada data pengujian. Metode ini digunakan untuk mengevaluasi kemampuan model dalam mengklasifikasikan kondisi aliran air berdasarkan empat komponen utama, yaitu *True Positive* (TP), *True Negative* (TN), *False positive* (FP), dan *False negative* (FN) [22].

Berdasarkan nilai yang diperoleh dari *confusion matrix*, dilakukan perhitungan empat metrik evaluasi utama, yaitu akurasi, presisi, *recall*, dan *F1-score*. Empat metrik utama tersebut dihitung sebagai berikut:

$$Akurasi = (TP + TN) / (TP + TN + FP + FN) \tag{5}$$

$$Presisi = TP / (TP + FP) \tag{6}$$

$$Recall = TP / (TP + FN) \tag{7}$$

$$F1 - Score = 2 \times (Presisi \times Recall) / (Presisi + Recall) \tag{8}$$

Dalam konteks penelitian ini, nilai presisi digunakan untuk mengukur kemampuan model dalam mengurangi alarm palsu (*false positive*), yaitu kondisi ketika sistem mendeteksi adanya aliran air padahal pipa sebenarnya kosong. Sementara itu, nilai *recall* digunakan untuk mengukur kemampuan model dalam mendeteksi seluruh kejadian aliran air yang sebenarnya sehingga meminimalkan kegagalan notifikasi (*false negative*). *F1-score* digunakan sebagai indikator keseimbangan antara kedua metrik tersebut sehingga dapat memberikan evaluasi yang lebih komprehensif terhadap performa model [22].

3. HASIL DAN PEMBAHASAN

3.1 Pengaruh Penyeimbangan Data Menggunakan SMOTE

Sebelum proses pelatihan model dilakukan, data latih yang diperoleh dari hasil pembagian *dataset* menunjukkan adanya ketidakseimbangan jumlah sampel antar kelas. Kondisi ini berpotensi menyebabkan model lebih dominan mempelajari pola dari kelas mayoritas sehingga kemampuan klasifikasi terhadap kelas minoritas menjadi kurang optimal. Untuk mengatasi permasalahan tersebut diterapkan metode *Synthetic Minority Over-sampling Technique* (SMOTE) pada data latih. Distribusi kelas sebelum dan sesudah penerapan SMOTE ditunjukkan pada Tabel 6.

Tabel 6. Distribusi Kelas Sebelum dan Sesudah SMOTE

Kelas	Sebelum SMOTE	Sesudah SMOTE
Air Mengalir (Label 1)	2.007	2.007
Pipa Kosong (Label 0)	1.517	2.007
Total	3.524	4.014

Berdasarkan Tabel 6 terlihat bahwa sebelum penerapan SMOTE jumlah sampel pada kelas pipa kosong lebih sedikit dibandingkan kelas air mengalir. Ketidakseimbangan tersebut berpotensi menyebabkan model

menghasilkan prediksi yang lebih condong terhadap kelas mayoritas. Setelah proses SMOTE dilakukan, jumlah sampel pada kedua kelas menjadi seimbang sehingga model memperoleh kesempatan yang sama untuk mempelajari karakteristik masing-masing kelas.

Penerapan SMOTE juga memberikan keuntungan dalam meningkatkan representasi data pada kelas minoritas tanpa harus mengumpulkan data baru secara langsung. Sampel sintetis yang dihasilkan membantu memperluas variasi data pelatihan sehingga model dapat membangun batas pemisah (*decision boundary*) yang lebih baik antara kedua kelas. Kondisi ini menjadi penting karena penelitian bertujuan untuk membedakan pola getaran yang terkadang memiliki karakteristik yang mirip antara kondisi air mengalir dan kondisi pipa kosong.

Selain itu, distribusi kelas yang lebih seimbang membantu mengurangi bias model terhadap kelas tertentu. Dengan demikian, performa klasifikasi yang diperoleh tidak hanya ditunjukkan oleh nilai akurasi yang tinggi, tetapi juga oleh kemampuan model dalam mengenali kedua kelas secara proporsional. Hal ini terlihat dari nilai *precision*, *recall*, dan *F1-score* yang tinggi pada kedua algoritma yang diuji.

3.2 Hasil Klasifikasi Menggunakan *Random Forest*

Model *Random Forest* dilatih menggunakan data hasil normalisasi dan penyeimbangan kelas menggunakan SMOTE. Parameter yang digunakan adalah $n_estimators = 100$ dan $random_state = 42$. Hasil evaluasi model *Random Forest* ditunjukkan pada Tabel 7.

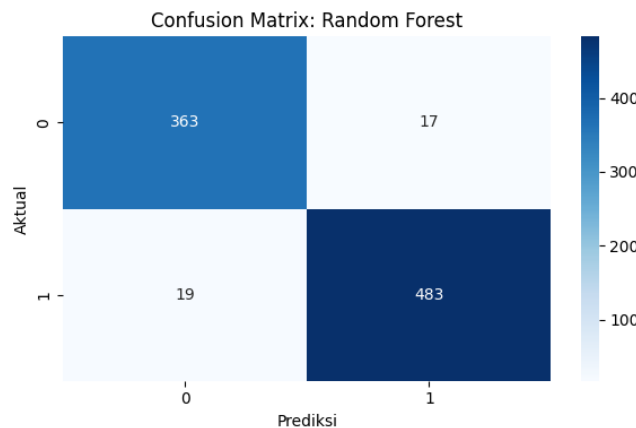
Tabel 7. Hasil Evaluasi *Random Forest*

Metrik	Nilai (%)
Accuracy	95,92
Precision	96,60
Recall	96,22
F1-score	96,41

Confusion matrix hasil pengujian model *Random Forest* ditunjukkan pada Tabel 8 dan Gambar 2.

Tabel 8. Confusion Matrix *Random Forest*

Aktual / Prediksi	Mengalir	Kosong	Total
Mengalir	TP = 483	FN = 19	502
Kosong	FP = 17	TN = 363	380
Total	500	382	882



Gambar 2. Visualisasi Hasil Confusion Matrix *Random Forest*

Berdasarkan *confusion matrix* pada Gambar 2, model *Random Forest* berhasil mengklasifikasikan 363 data pipa kosong (*True Negative*) dan 483 data air mengalir (*True Positive*) secara tepat. Dari total 882 data uji, model hanya menghasilkan 36 kesalahan klasifikasi. Kesalahan tersebut secara spesifik terdiri atas 17 data pipa kosong yang salah diprediksi sebagai air mengalir (*false positive*) dan 19 data air mengalir yang gagal terdeteksi dan diprediksi sebagai pipa kosong (*false negative*). Tingkat kesalahan yang sangat minim ini menjadi bukti langsung bahwa model *Random Forest* mampu memisahkan karakteristik getaran antara pipa beraliran dan pipa kosong secara efektif.

Ditinjau dari rincian metrik evaluasinya, tingginya nilai presisi sebesar 96,60% menunjukkan bahwa dari seluruh prediksi "air mengalir" yang dikeluarkan oleh model, hampir keseluruhannya tepat sasaran dan merupakan kejadian aliran air yang sebenarnya. Secara praktis, presisi tinggi ini sangat esensial karena merepresentasikan rasio alarm palsu (*false alarm*) yang sangat rendah pada sistem. Di sisi lain, nilai *recall* sebesar 96,22% menegaskan tingginya sensitivitas model. Angka tersebut membuktikan bahwa model sukses mendeteksi sebagian

besar kondisi riil saat air benar-benar mengalir, dengan jumlah kejadian yang terlewatkan (*false negative*) di bawah angka 20 sampel.

Kombinasi performa presisi dan *recall* yang sama-sama kuat ini secara matematis menghasilkan nilai *F1-score* sebesar 96,41%. Pencapaian *F1-score* ini mengonfirmasi bahwa performa model tetap stabil dan tidak bias terhadap salah satu kelas pasca-penerapan SMOTE. Secara keseluruhan, dengan tingkat akurasi mencapai 95,92%, hasil pengujian ini menyimpulkan bahwa algoritma *Random Forest* sangat andal dalam menginterpretasikan fluktuasi fitur akselerasi dan statistik dari sensor MPU6050 menjadi informasi kondisi aliran air yang presisi.

3.3 Hasil Klasifikasi Menggunakan *Support Vector Machine*

Model *Support Vector Machine* (SVM) dilatih menggunakan kernel *Radial Basis Function* (RBF). Sebelum proses pelatihan dilakukan normalisasi data menggunakan *StandardScaler* serta penyeimbangan kelas menggunakan SMOTE. Hasil evaluasi model SVM ditunjukkan pada Tabel 9.

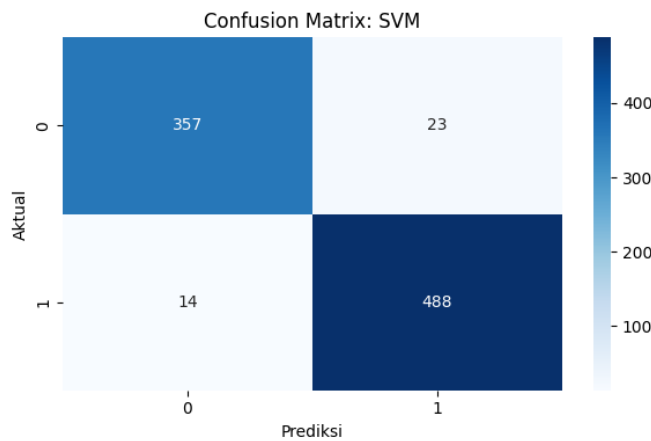
Tabel 9. Hasil Evaluasi *Support Vector Machine*

Metrik	Nilai (%)
Accuracy	95,80
Precision	95,50
Recall	97,21
<i>F1-score</i>	96,35

Confusion matrix hasil pengujian model SVM ditunjukkan pada Tabel 10 dan Gambar 3.

Tabel 10. Confusion Matrix Algoritma *Support Vector Machine*

Aktual / Prediksi	Mengalir	Kosong	Total
Mengalir	TP = 488	FN = 14	502
Kosong	FP = 23	TN = 357	380
Total	511	371	882



Gambar 3. Visualisasi Hasil Confusion Matrix Algoritma *Support Vector Machine*

Berdasarkan *confusion matrix* pada Gambar 3, model SVM berhasil mengklasifikasikan 357 data pipa kosong (*True Negative*) dan 488 data air mengalir (*True Positive*) secara tepat. Dari keseluruhan 882 sampel data uji, model menghasilkan total 37 kesalahan klasifikasi. Kesalahan ini terdistribusi menjadi 23 data pipa kosong yang diprediksi keliru sebagai air mengalir (*false positive*) dan 14 data air mengalir yang gagal terdeteksi sehingga diklasifikasikan sebagai pipa kosong (*false negative*). Tingkat akurasi keseluruhan sebesar 95,80% menegaskan bahwa fitur getaran yang digunakan memiliki kemampuan diskriminatif yang baik sehingga dapat dipisahkan secara efektif oleh *hyperplane* yang dibentuk oleh model SVM, bahkan dalam kondisi hubungan *non-linear* sekalipun berkat penggunaan *kernel RBF*.

Ditinjau dari rincian metrik evaluasinya, tingginya nilai *recall* sebesar 97,21% menjadi keunggulan absolut dari algoritma SVM. Angka *recall* yang sangat tinggi ini membuktikan tingginya sensitivitas model dalam mendeteksi kejadian riil saat air benar-benar mengalir, dengan jumlah kegagalan deteksi (*false negative*) yang sangat minim (hanya 14 sampel). Dalam konteks pemantauan distribusi air, sensitivitas tinggi ini merupakan aspek yang sangat krusial karena memastikan sistem peringatan dini hampir tidak pernah melewatkan momen ketersediaan debit air sekecil apa pun, sehingga masyarakat tidak kehilangan kesempatan untuk menampung air.

Namun demikian, peningkatan kepekaan deteksi tersebut diikuti dengan bertambahnya jumlah *false positive* (23 sampel). Hal ini mengindikasikan bahwa SVM cenderung sedikit lebih agresif dalam memberikan prediksi positif, yang berimbas pada sedikit menurunnya nilai presisi menjadi 95,50%. Terdapat *trade-off* berupa sedikit peningkatan risiko alarm palsu (*false alarm*) dibandingkan model *Random Forest*.

Meskipun demikian, perbedaan performa ini tergolong sangat kecil. Kombinasi presisi dan *recall* pada SVM tetap menghasilkan keseimbangan yang sangat solid dengan perolehan nilai *F1-score* sebesar 96,35%. Hasil pengujian ini mengonfirmasi bahwa algoritma SVM memiliki kapabilitas yang sangat andal dalam menginterpretasikan fluktuasi sinyal sensor menjadi informasi kondisi aliran air yang akurat pada *dataset* yang digunakan.

3.4 Perbandingan Performa Model dan Analisis Kesalahan Klasifikasi

Tabel 11 menyajikan perbandingan seluruh metrik evaluasi antara *Random Forest* dan SVM.

Tabel 11. Komparasi Metrik Evaluasi *Random Forest* vs. SVM

Model	Akurasi (%)	Presisi (%)	Recall (%)	F1-score (%)
<i>Random Forest</i>	95,92	96,60	96,22	96,41
<i>Support Vector Machine</i>	95,80	95,50	97,21	96,35

Berdasarkan Tabel 11 terlihat bahwa kedua algoritma mampu menghasilkan performa klasifikasi yang sangat baik dengan nilai akurasi di atas 95%. *Random Forest* memperoleh nilai akurasi tertinggi sebesar 95,92%, sedangkan SVM memperoleh nilai akurasi sebesar 95,80%. Selisih akurasi antara kedua model hanya sebesar 0,12%, yang menunjukkan bahwa keduanya memiliki kemampuan yang hampir setara dalam membedakan kondisi air mengalir dan pipa kosong berdasarkan data getaran yang digunakan.

Pada metrik presisi, *Random Forest* memperoleh nilai sebesar 96,60%, lebih tinggi dibandingkan SVM yang memperoleh nilai 95,50%. Hasil tersebut menunjukkan bahwa *Random Forest* lebih baik dalam mengurangi prediksi positif yang salah (*false positive*). Dalam konteks sistem pemantauan distribusi air, nilai presisi yang tinggi penting karena dapat mengurangi kemungkinan sistem memberikan informasi bahwa air sedang mengalir padahal kondisi pipa sebenarnya kosong. Sebaliknya, SVM memperoleh nilai *recall* yang lebih tinggi, yaitu 97,21%, dibandingkan *Random Forest* sebesar 96,22%. Nilai *recall* yang tinggi menunjukkan bahwa SVM lebih mampu mendeteksi kondisi air mengalir yang sebenarnya sehingga mengurangi kemungkinan terjadinya kegagalan deteksi.

Untuk memahami lebih lanjut perbedaan performa kedua model, dilakukan analisis terhadap kesalahan klasifikasi yang dihasilkan selama proses pengujian. Perbandingan jumlah *false positive* (FP) dan *false negative* (FN) pada kedua model ditunjukkan pada Tabel 12.

Tabel 12. Perbandingan Kesalahan Klasifikasi Model

Model	False positive (FP)	False negative (FN)	Total Kesalahan
<i>Random Forest</i>	17	19	36
<i>Support Vector Machine</i>	23	14	37

Berdasarkan Tabel 12 terlihat bahwa jumlah total kesalahan yang dihasilkan kedua model relatif berdekatan. *Random Forest* menghasilkan 36 kesalahan klasifikasi, sedangkan SVM menghasilkan 37 kesalahan klasifikasi. Meskipun demikian, distribusi jenis kesalahan yang dihasilkan menunjukkan karakteristik yang berbeda pada masing-masing algoritma.

Random Forest menghasilkan jumlah *false positive* yang lebih sedikit dibandingkan SVM. Hal ini menunjukkan bahwa *Random Forest* cenderung lebih konservatif dalam memberikan prediksi kondisi air mengalir. Dengan kata lain, model hanya memberikan prediksi positif ketika memiliki tingkat keyakinan yang cukup tinggi terhadap pola data yang diamati. Karakteristik tersebut berdampak pada meningkatnya nilai presisi karena jumlah alarm yang tidak diperlukan dapat dikurangi. Pada implementasi nyata, kondisi ini mengurangi kemungkinan masyarakat menerima informasi bahwa air telah mengalir ketika pada kenyataannya pipa masih dalam kondisi kosong.

Di sisi lain, SVM menghasilkan jumlah *false negative* yang lebih rendah dibandingkan *Random Forest*. Kesalahan *false negative* terjadi ketika kondisi air mengalir diklasifikasikan sebagai pipa kosong. Dalam sistem pemantauan distribusi air, kesalahan jenis ini dapat menyebabkan pengguna tidak memperoleh informasi bahwa air sedang tersedia sehingga berpotensi kehilangan kesempatan untuk melakukan penampungan air. Jumlah *false negative* yang lebih sedikit pada SVM menunjukkan bahwa model memiliki sensitivitas yang lebih tinggi dalam mengenali keberadaan aliran air. Temuan ini juga sejalan dengan nilai *recall* SVM yang lebih tinggi dibandingkan *Random Forest*.

Perbedaan karakteristik kesalahan tersebut menunjukkan adanya *trade-off* antara kemampuan mengurangi alarm palsu dan kemampuan mendeteksi seluruh kejadian positif. *Random Forest* lebih unggul dalam mengurangi *false positive*, sedangkan SVM lebih unggul dalam mengurangi *false negative*. Oleh karena itu, pemilihan model terbaik dapat disesuaikan dengan kebutuhan implementasi sistem. Jika sistem lebih memprioritaskan pengurangan alarm yang tidak diperlukan, maka *Random Forest* menjadi pilihan yang lebih sesuai. Sebaliknya, apabila prioritas utama adalah memastikan bahwa setiap kejadian aliran air dapat terdeteksi, maka SVM dapat menjadi alternatif yang lebih baik.

Meskipun demikian, berdasarkan evaluasi keseluruhan yang mencakup akurasi, presisi, *recall*, dan *F1-score*, *Random Forest* menunjukkan performa yang sedikit lebih unggul dibandingkan SVM. Nilai *F1-score* sebesar 96,41% menunjukkan bahwa *Random Forest* mampu menjaga keseimbangan yang baik antara presisi dan *recall*. Selain itu, jumlah total kesalahan yang lebih rendah serta nilai presisi yang lebih tinggi menunjukkan bahwa model ini lebih konsisten dalam menghasilkan prediksi yang akurat pada *dataset* yang digunakan. Oleh karena itu, *Random Forest* dipilih sebagai model terbaik pada penelitian ini untuk klasifikasi kondisi aliran air berdasarkan data getaran pipa distribusi.

3.5 Pembahasan

Hasil penelitian menunjukkan bahwa data getaran yang diperoleh dari sensor MPU6050 mampu digunakan sebagai indikator yang efektif untuk membedakan kondisi air mengalir dan kondisi pipa kosong. Ketika air mengalir di dalam pipa, interaksi antara fluida dan dinding pipa menghasilkan pola getaran yang berbeda dibandingkan kondisi pipa tanpa aliran. Perbedaan pola tersebut terekam oleh sensor akselerometer dan direpresentasikan melalui fitur akselerasi tiga sumbu serta fitur statistik berupa *mean* dan RMS. Nilai RMS secara khusus berperan dalam menggambarkan energi getaran yang dihasilkan selama proses pengukuran, sedangkan nilai *mean* memberikan informasi mengenai kecenderungan amplitudo sinyal dalam suatu periode pengamatan.

Selain dipengaruhi oleh kualitas fitur, performa model juga didukung oleh tahapan pra-pemrosesan data yang dilakukan sebelum proses pelatihan. Normalisasi menggunakan *StandardScaler* membantu menyamakan rentang nilai antar fitur sehingga proses pembelajaran model menjadi lebih optimal, terutama pada algoritma SVM yang sensitif terhadap perbedaan skala data. Di sisi lain, penerapan SMOTE berhasil mengatasi ketidakseimbangan kelas pada data latih sehingga model tidak terlalu bias terhadap kelas mayoritas. Kombinasi kedua tahapan tersebut berkontribusi terhadap tingginya nilai akurasi yang diperoleh oleh kedua algoritma.

Jika dibandingkan dengan penelitian sebelumnya, hasil penelitian ini menawarkan pembuktian empiris mengenai efektivitas dan kepraktisan pendekatan berbasis getaran (*vibration-based*) dibandingkan metode akustik yang umum digunakan. Sebagai perbandingan, penelitian sebelumnya oleh Liu et al. [4] dan Pourmehrani et al. [6] menitikberatkan pada penggunaan sinyal vibroakustik dan suara untuk mendeteksi kebocoran, yang sering kali rentan terhadap gangguan *noise* lingkungan di udara terbuka. Berbeda dengan pendekatan mereka, penelitian ini menunjukkan bahwa pembacaan getaran fisik secara langsung melalui sensor akselerometer *non-intrusif* (MPU6050) pada permukaan pipa mampu memberikan ketahanan pembacaan data yang memadai untuk diklasifikasikan oleh model komputasi secara *real-time*. Lebih lanjut, studi literatur oleh Rajgopal et al. [13] serta Widyatma et al. [5] telah memvalidasi betapa arsitektur *Wireless Sensor Network* (WSN) dan IoT dapat memodernisasi infrastruktur pemantauan air. Namun, penelitian-penelitian tersebut mayoritas berfokus pada topologi pengiriman datanya saja atau pembacaan sensor aliran air (*water flow sensor*) yang mengharuskan pemotongan pipa fisik (*intrusif*). Sebagai bentuk kebaruan (*novelty*), penelitian ini mengisi *research gap* tersebut dengan membuktikan bahwa arsitektur integrasi data sensor *non-intrusif* dan algoritma klasifikasi *machine learning* mampu menyelesaikan masalah identifikasi jadwal kedatangan air tanpa perlu memodifikasi infrastruktur perpipaan warga yang sudah ada.

Pada konteks Desa Lubuk Raman, sistem yang dikembangkan berpotensi membantu masyarakat memperoleh informasi mengenai kondisi aliran air tanpa perlu melakukan pengecekan secara manual. Dengan mengintegrasikan model klasifikasi ke dalam perangkat IoT, status aliran air dapat dipantau secara otomatis dan dikirimkan melalui media notifikasi kepada pengguna. Informasi tersebut dapat membantu masyarakat mempersiapkan penampungan air pada waktu yang tepat serta mengurangi ketergantungan terhadap pengecekan manual yang selama ini dilakukan.

Meskipun menghasilkan performa yang tinggi, penelitian ini masih memiliki beberapa keterbatasan. *dataset* yang digunakan diperoleh dari satu lokasi pengujian sehingga variasi kondisi lingkungan yang berbeda belum sepenuhnya terwakili. Selain itu, penelitian hanya membandingkan dua algoritma klasifikasi, yaitu *Random Forest* dan SVM. Penelitian selanjutnya dapat dilakukan dengan memperluas jumlah data, menambahkan fitur ekstraksi sinyal yang lebih kompleks, serta membandingkan metode *machine learning* maupun deep learning lainnya untuk memperoleh performa yang lebih optimal.

4. KESIMPULAN

Pemanfaatan data getaran *non-intrusif* yang diperoleh secara langsung dari permukaan pipa menggunakan sensor MPU6050 telah menunjukkan potensi yang sangat kuat untuk diimplementasikan sebagai indikator utama keberadaan aliran air pada sistem distribusi. Karakteristik percepatan pada tiga sumbu pengukuran (*ax*, *ay*, *az*) yang dipadukan dengan parameter statistik terukur berupa *mean* dan *Root Mean Square* (RMS) terbukti mampu membentuk representasi data yang diskriminatif untuk membedakan secara tegas antara kondisi pipa beraliran dan pipa kosong. Hasil evaluasi eksperimental memperlihatkan bahwa kedua algoritma *machine learning* yang diuji mampu mencapai tingkat klasifikasi yang luar biasa tinggi dengan margin perbedaan performa yang relatif kecil. Berdasarkan hasil uji coba terhadap 882 sampel data pengujian, algoritma *Random Forest* berhasil mendapatkan skor akurasi 95,92%, presisi 96,60%, *recall* 96,22%, dan *F1-score* 96,41%. Sementara itu, algoritma *Support*

Vector Machine (SVM) mencapai skor akurasi 95,80%, presisi 95,50%, *recall* 97,21%, dan *F1-score* 96,35%. Pencapaian komprehensif tersebut secara meyakinkan menunjukkan bahwa kedua algoritma memiliki kapabilitas performa yang sangat baik dengan tingkat akurasi di atas 95 persen. Namun demikian, *Random Forest* terbukti sedikit lebih unggul secara holistik dibandingkan SVM, terutama dalam kemampuannya menekan angka *false positive* yang direpresentasikan oleh tingginya nilai presisi dan stabilitas *F1-score*. Pemilihan *Random Forest* sebagai model terbaik menjadikannya sangat ideal untuk meminimalisasi alarm palsu dalam operasional sistem harian. Dengan demikian, tujuan utama penelitian ini untuk menganalisis serta membandingkan secara komparatif kinerja algoritma *Random Forest* dan SVM dalam mengklasifikasikan kondisi aliran air berbasis pengenalan pola getaran telah berhasil dicapai sepenuhnya. Meskipun memberikan hasil yang menjanjikan, penelitian ini masih memiliki sejumlah batasan empiris, terutama terkait jumlah sampel dan keragaman variasi gangguan lingkungan yang difokuskan pada satu lokasi observasi, serta ruang lingkup komparasi yang hanya melibatkan dua metode klasifikasi. Oleh karena itu, penelitian selanjutnya disarankan untuk memperluas variasi kondisi perekaman data di berbagai topologi perpipaan yang berbeda dan mengeksplorasi ekstraksi fitur sinyal yang lebih kompleks.

REFERENCES

- [1] LOCALISE SDGs Indonesia, "Sustainable Development Goals." Accessed: Feb. 16, 2026. [Online]. Available: <https://localisesdgs-indonesia.org/17-sdgs>
- [2] Badan Pusat Statistik RI, "Statistik Lingkungan Hidup Indonesia 2024," Jakarta, 2024.
- [3] Z. Liu and W. Liu, "Spatial-Temporal Relationship between Water Resources and Economic Development in Rural China from a Poverty Perspective," *Int. J. Environ. Res. Public Heal.*, 2021, doi: 10.3390/ijerph18041540.
- [4] R. Liu, T. Zayed, and R. Xiao, "A critical systematic review of machine learning based models for water leak detection using vibroacoustic technology," 2025, doi: 10.1016/j.engappai.2025.111432.
- [5] D. H. Widyatma *et al.*, "Internet of things-based water pipe leak detection system," 2025, doi: 10.26623/elektrika.v17i2.12750.
- [6] H. Pourmehrani, R. Hosseini, and H. Moradi, "Water Flow Detection Device Based on Sound Data Analysis and Machine Learning to Detect Water Leakage," 2025.
- [7] S. A. Palermo *et al.*, "Smart Technologies for Water Resource Management : An Overview," *Sensors*, 2022, doi: 10.3390/s22166225.
- [8] M. Zuñiga-uribe, R. Rojas-galván, J. M. Álvarez-alvarado, M. Aviles, G. I. Pérez-soto, and V. Pérez-moreno, "Artificial Intelligence in Water Distribution Networks : A Systematic Review of Models , Input Variables , Databases , and Output Strategies for Leak Detection," pp. 1–29, 2026.
- [9] M. Dada, M. Majemite, A. Obaigbena, O. Daraojimba, J. Oliha, and Z. Nwokediegwu, "Review of smart water management: IoT and AI in water and wastewater treatment," *World J. Adv. Res. Rev.*, 2024, doi: 10.30574/wjarr.2024.21.1.0171.
- [10] G. Dharmarathne, A. M. S. R. Abekoon, M. Bogahawaththa, J. Alawatugoda, and P. Meddage, "A review of machine learning and internet-of-things on the water quality assessment: Methods, applications and future trends," *Results Eng.*, 2025, doi: 10.1016/j.rineng.2025.105182.
- [11] H. Cen, D. Huang, Q. Liu, Z. Zong, and A. Tang, "Application Research on Risk Assessment of Municipal Pipeline Network Based on *Random Forest* Machine Learning Algorithm," *Water*, 2023, doi: 10.3390/w15101964.
- [12] J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua, and A. Lopez, "A comprehensive survey on support vector machine classification: Applications," *Neurocomputing*, vol. 408, pp. 189–215, 2020, doi: 10.1016/j.neucom.2019.10.118.
- [13] M. Rajgopal, K. P. Karani, and N. S. Basavaraju, "Cluster based water leakage detection frame work for the improvement of water management using WSN," vol. 38, no. 1, pp. 603–612, 2025, doi: 10.11591/ijeecs.v38.i1.pp603-612.
- [14] A. Pagano, D. Garlisi, I. Tinnirello, and F. Giuliano, "Ad Hoc Networks Survey paper A survey on massive IoT for water distribution systems : Challenges , simulation tools , and guidelines for large-scale deployment," *Ad Hoc Networks*, vol. 168, no. October 2024, p. 103714, 2025, doi: 10.1016/j.adhoc.2024.103714.
- [15] I. H. Sarker, "Machine Learning: Algorithms, Real-World Applications and Research Directions," *SN Comput. Sci.*, vol. 2, 2021, doi: 10.1007/s42979-021-00592-x.
- [16] S. Lee and B. Kim, "Machine Learning Model for Leak Detection Using Water Pipeline Vibration Sensor.," *Sensors*, 2023, doi: 10.3390/s23218935.
- [17] J. M. Johnson and T. M. Khoshgoftaar, "Survey on deep learning with class imbalance.," *J Big Data*, 2019, doi: 10.1186/s40537-019-0192-5.
- [18] J. Herrera Pinheiro *et al.*, "The Impact of Feature Scaling In Machine Learning: Effects on Regression and Classification Tasks," 2025, doi: 10.1109/ACCESS.2025.3635541.
- [19] A. Géron, *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow*, 2nd ed. Sebastopol:



- O'Reilly Media, Inc., 2019.
- [20] S. Wang, Y. Dai, J. Shen, and J. Xuan, "Research on expansion and classification of imbalanced data based on SMOTE algorithm," *Sci. Rep.*, 2021, doi: 10.1038/s41598-021-03430-5.
- [21] J. Han, M. Kamber, and J. Pei, *Data mining: concepts and techniques*, 3rd ed. Waltham: Morgan Kaufmann, 2012.
- [22] D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," *BMC Genomics*, 2020, doi: 10.1186/s12864-019-6413-7.