

RANCANG BANGUN SISTEM PREDIKSI USIA BERJALAN DARI GAYA BERJALAN (*GAIT*) MANUSIA MENGGUNAKAN METODE *K-NEAREST NEIGHBORS*

Hafidh Al Asad¹, Husneni Mukhtar², Dien Rahmawati³,
^{1,2,3}Prodi S1 Teknik Elektro, Fakultas Teknik Elektro, Universitas Telkom

¹hafidhalasad@student.telkomuniversity.ac.id, ²husnenimukhtar@telkomuniversity.ac.id,
³dienrahmawati@telkomuniversity.ac.id,

Diterima pada 17 September 2022; disetujui pada 8 November 2022; dan diterbitkan pada 31 Desember 2022.

Abstrak

Penelitian terkait gaya berjalan manusia (*human gait*) masih terus menarik minat para peneliti sampai saat ini, di mana manfaatnya tidak lagi hanya melihat karakteristik dan indikasi fisiologi pada manusia, namun juga telah merambah ke era digitalisasi berteknologi tinggi dengan melakukan analisis gaya berjalan untuk berbagai aplikasi seperti biometrik, otentikasi, dan berbagai keperluan analisis lainnya. Analisis gaya berjalan untuk memprediksi usia manusia adalah salah satu pengembangan riset *human gait*, terutama menggunakan sistem instrumentasi dan pembelajaran mesin. Penelitian ini menggunakan sensor inersia dan metode *K-Nearest Neighbors* (K-NN) dengan ekstraksi fitur *mean* dalam memprediksi usia berjalan seseorang dari gaya berjalannya. Sebanyak 25 partisipan telah direkam data kecepatan berjalannya dengan tiga kelompok usia, maka prediksi usia seseorang telah berhasil diperoleh dengan akurasi sebesar 86,9% dengan menggunakan 80% data latih dan 20% data uji, sedangkan metode pengklasifikasi K-NN dipilih berdasarkan hasil nilai *cross validation* (CV) terbaik dibandingkan metode lainnya.

Kata Kunci: walking age, K-Nearest Neighbors (KNN), gaya berjalan, sensor inersia, klasifikasi

Abstract

Research related to human gait continues to attract the interest of researchers, where the benefits are not only looking at the characteristics and indications of human physiology, but they have also developed forward to the era of high digitalization technology by conducting gait analysis for various applications such as biometrics, authentication, and various other analytical purposes. Gait analysis to predict human age is one of the developments of human gait research, especially using instrumentation systems and machine learning. This study uses two inertial sensors and K-NN method with the feature extraction of mean in predicting a person's walking age from his gait. As many as 25 participants have recorded walking speed data with three age groups, then the prediction of a person's age has been successfully obtained with an accuracy of 86.9% using 80% training data and 20% test data, while the K-NN classification method was chosen based on the results of the cross validation (CV) value is the best one compared to other methods.

Key Words: walking age, K-Nearest Neighbors (KNN), gait, inertial sensor, classification.

1. Pendahuluan

Gaya berjalan (*gait*) merupakan suatu gerakan yang melibatkan koordinasi saraf pada otak dan sistem lainnya di tubuh manusia yang sangat rumit. Setiap manusia memiliki otot, otak, sistem saraf, dan postur tubuh yang berbeda, apalagi dengan adanya penambahan usia. Hal ini menjadikan hampir setiap manusia mempunyai gaya berjalan yang berbeda-beda [1].

Penggunaan sensor inersia (akselerometer dan giroskop) sangatlah populer untuk penelitian *gait*

analysis atau analisis gaya berjalan manusia karena sensor ini menawarkan pengukuran yang mudah, terjangkau, dan dapat dikenakan atau disematkan pada perangkat yang kita gunakan dalam kehidupan sehari-hari, seperti jam tangan, gawai, dan sebagainya. Keuntungan riset analisis gaya berjalan ini sangatlah luas dan besar, di mana gaya berjalan merupakan fitur biometrik penting yang dapat membantu mengidentifikasi manusia dan tindakannya. Selain itu, analisis gaya berjalan dapat digunakan dalam berbagai

aplikasi termasuk otentikasi, keamanan, biomekanik, olahraga, dan bidang medis.

Maraknya perkembangan kebutuhan biometrik, terutama dalam menentukan dan memprediksi usia seseorang, menjadikan penggunaan sensor inersia ini dilirik sebagai alternatif solusi dari ciri gaya berjalan seseorang. Terlebih lagi, estimasi usia menggunakan metode pengenalan wajah tidak dapat diandalkan akurasi karena wajah yang terlihat muda belum tentu merepresentasikan usia yang muda juga, begitupun sebaliknya. Akurasi yang diperoleh sebaiknya menggunakan teknik evaluasi metrik dari fitur yang kontras atau berbeda antar kelompoknya sehingga pengklasifikasian dapat berjalan dengan optimal [2].

Kemudian, mengingat bahwa gaya berjalan sangat sensitif terhadap penuaan, maka analisis gaya berjalan merupakan salah satu solusi yang cocok untuk mengestimasi usia seseorang [3]. Kegunaan dari estimasi usia berjalan ini sangat dibutuhkan dalam analisis di beberapa bidang, seperti bidang pemasaran, yaitu dalam proses penghitungan pelanggan otomatis untuk riset pemasaran atau pembatasan akses kontrol berbasis usia otomatis untuk area tertentu. Selain menggunakan sensor yang dikenakan, gaya berjalan dapat diamati dari jarak jauh menggunakan kamera, seperti CCTV pada area tertentu [4].

Prediksi usia manusia dari cara berjalan, dalam suatu riset, dilakukan menggunakan sensor wltlt v3.0 dengan cara merekam gerak jalan partisipan yang usianya bervariasi menggunakan metode *K-Nearest Neighbors* (K-NN) [5] untuk mendapatkan informasi kesehatan seseorang melalui cara berjalannya. Namun, akurasi yang diperoleh terbilang kurang memuaskan. Jumlah partisipan yang besar, yaitu 79 partisipan, dengan teknologi sensor yang baik hanya mampu menghasilkan akurasi sebesar 87%, itupun setelah dilakukan penyesuaian pengecualian terhadap kelompok usia tertentu agar akurasi dapat meningkat.

Pada penelitian ini, diusulkan suatu eksperimen pengukuran gaya berjalan manusia untuk memprediksi usia seseorang menggunakan sistem klasifikasi yang mudah dan sederhana, yaitu pembacaan kecepatan berjalan dalam 3 dimensi dengan sensor inersia dan metode pengklasifikasian *K-Nearest Neighbours* (K-NN). Metode K-NN yang digunakan pada kasus ini merupakan metode dengan hasil akurasi terbaik dibandingkan beberapa metode lainnya, seperti AdaBoost (AB), GradientBoost (GB), *random forest* (RF), *decision tree* (DT), *Gaussian Naïve Bayes* (NB), dan *support vector machine* (SVM).

2. Metode

Subbagian ini berisikan metode yang digunakan dalam penelitian ini, meliputi pra-pengolahan data gaya berjalan yang diakusisi dari sensor akselerometer,

ekstraksi fitur yang digunakan, dan beberapa metode pembelajaran mesin yang digunakan dan selanjutnya dipilih berdasarkan nilai akurasi pengujian.

2.1 Pra pengolahan data

Keberhasilan pembelajaran mesin dapat ditentukan dengan melakukan langkah pengolahan data awal (*pre-processing*), yaitu menyiapkan data terlebih dahulu sebelum data dilatih. *Pre-processing* itu sendiri dapat berupa pembersihan data, transformasi data, dan ekstraksi fitur. *Fast Fourier Transform* (FFT) adalah teknik *pre-processing* yang digunakan dalam penelitian ini, yaitu melakukan operasi yang mengubah sinyal dari domain waktu ke domain frekuensi menggunakan persamaan 1 [6]. Hasil *pre-processing* dilanjutkan ke tahap ekstraksi fitur.

$$X_k = \sum_{n=0}^{N-1} x_n e^{-2\pi i k n / N} \quad (1)$$

X_k = *Discrete Fourier Transform*

N = bilangan kompleks

X_n = nilai eksponensial

e^{i2}

$2\pi i k n / N$ = rumus Euler

2.2 Ekstraksi Fitur

Pengambilan objek berdasarkan fitur histogram yang merepresentasikan probabilitas kemunculan nilai yang dihasilkan adalah metode dari ekstraksi fitur [7], di mana metode ini bertujuan untuk mengurangi jumlah fitur dalam kumpulan data dengan membuat fitur baru dari yang sudah ada. Kumpulan fitur baru ini dapat merangkum sebagian besar informasi yang terkandung dalam kumpulan fitur asli yang dihitung dari beberapa parameter seperti *mean*, *median*, maksimum, minimum, *skew*, *kurtosis*, dan *entropi*. *Mean* atau rata-rata merupakan ukuran total nilai sampel dibagi dengan jumlah data. *Median* merupakan nilai tengah dari kumpulan data, *max* atau nilai maksimum merupakan nilai sampel tertinggi dari data yang ada pada sekumpulan data, sedangkan *min* atau nilai minimum merupakan nilai sampel terkecil dari data yang ada pada sekumpulan data. *Skew* merupakan tingkat *k* relatif kurva dari kumpulan data dan *kurtosis* merupakan derajat kepuncakan relatif dari suatu distribusi data normal.

2.3 Classifier

Penelitian ini menggunakan beberapa metode *classifier* pada pembelajaran mesin yang umum digunakan dalam pengolahan data sensor untuk keperluan rekognisi atau identifikasi aktivitas manusia, seperti klasifikasi AdaBoost (AB), GradientBoost (GB), *random forest* (RF), *K-Nearest Neighbors* (KNN),

decision tree (DT), *Gaussian Naïve Bayes* (NB) dan *support vector machine* (SVM).

AdaBoost dan GradientBoost termasuk ke dalam algoritma *boosting*, yaitu metode ensemble dalam meningkatkan performansi suatu model dengan mengurangi kejadian bias dan varians [8]. Klasifikasi dengan AdaBoost (*Adaptive Boosting*) digunakan dalam permasalahan regresi, baik klasifikasi biner maupun klasifikasi *multiclass* tanpa pengecualian. Formula AdaBoost ditulis dalam persamaan 2 berikut.

$$H(x) = \text{sign} \left(\sum_{t=1}^T a_t h_t(x) \right) \quad (2)$$

di mana h_t adalah output dari pengklasifikasi t (nilai -1 atau +1), T adalah pengklasifikasi lemah yang digunakan, a_t adalah bobot yang digunakan dalam pengklasifikasi t . Dengan demikian, hasil akhirnya adalah hanya kombinasi linear dari seluruh pengklasifikasi lemah.

Prosedur pembelajaran pada *Gradient Boosting* mengungkap konsep pembelajaran model regresi adaptif yang secara berurutan menggunakan fungsi parametrik sederhana yang multivariable. Dengan kata lain, *gradient boosting* dikembangkan dari model *decision tree* karena metode ini terdiri dari banyak pohon klasifikasi yang secara bertahap dikembangkan dari pohon sebelumnya.

Support vector machine (SVM) merupakan sistem yang menggunakan fungsi hipotesis linear dengan fitur berdimensi tinggi dan dilatih menggunakan algoritma pembelajaran optimasi dengan bias pembelajaran statistik. Jenis sistem pembelajaran yang terawasi ini dalam melakukan klasifikasi dan regresinya berusaha mencari *hyperplanes* dengan tingkat pemisahan maksimum dan pengelompokan data dengan bantuan *support vector* sebagai referensi margin maksimum. Referensi tersebut dimodelkan dalam persamaan 3 dan berfungsi untuk memisahkan dan mengklasifikasikan data menggunakan *hyperplanes* [9], [10].

$$\arg \max \left\{ \frac{1}{w} \min [t_n (w^T k(x, x') + b)] \right\} \quad (3)$$

Random Forest (RF) merupakan kumpulan atau ansambel dari klasifikasi dan pohon regresi, atau terdiri dari satu set pengklasifikasi yang dilatih secara individual (seperti jaringan saraf atau pohon keputusan) yang prediksinya digabungkan untuk mengklasifikasikan kejadian baru. Dengan demikian, RF adalah *classifier* yang terdiri dari kumpulan pengklasifikasi berstruktur pohon $\{h(x, \Theta_k) = 1, 2, \dots\}$ di mana $\{\Theta_k\}$ adalah vektor acak terdistribusi identik yang independen dan setiap pohon membentuk unit yang memilih kelas paling populer pada input x

K-NN merupakan algoritma sederhana yang menyimpan semua data yang tersedia dan

mengklasifikasikan data baru berdasarkan ukuran kesamaan. Selain dikenal dengan implementasinya yang mudah dan sederhana, algoritma metode K-NN juga dapat memberikan kinerja yang relatif lebih tinggi untuk menghasilkan suatu model klasifikasi. Pendekatan klasifikasi data dilakukan dengan cara memperkirakan seberapa besar kemungkinan suatu titik data menjadi anggota dari suatu grup yang dilihat dari titik terdekat dari suatu grup [11]. Dalam menentukan dekat-jauhnya jarak antar tetangganya, metode ini menggunakan persamaan *Euclidean Distance* (persamaan 4).

$$D(x, y) = \sqrt{\sum_{i=1}^d (a_k - b_k)^2} \quad (4)$$

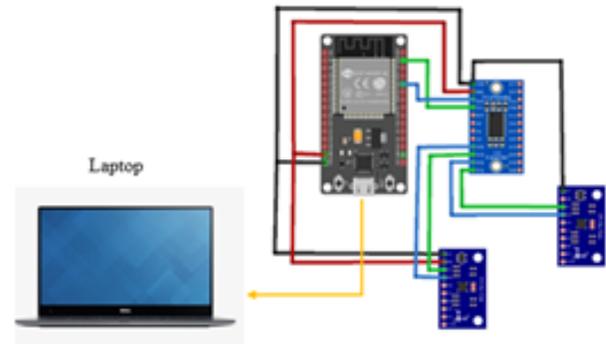
di mana $D(x, y)$ adalah jarak antara data uji dan data latih, a_k adalah data latih, b_k adalah data uji, dan d adalah dimensi data. Nilai k pada K-NN ditentukan dengan cara menghitung akar kuadrat dari jumlah total data yang dilatih [12]. Penentuan nilai k perlu dilakukan dengan cermat karena nilai k pada algoritma K-NN berhubungan dengan tingkat kesalahan model. Jika nilai k terlalu kecil maka hal tersebut dapat membuat algoritma lebih sensitif terhadap *noisy* data, walaupun hal tersebut dapat diatasi dengan menggunakan nilai k yang sedikit lebih besar. Namun, penentuan nilai k yang terlalu besar malah dapat menurunkan performansi sistem karena batasan antar tetangganya menjadi makin tidak jelas. Dengan kata lain, nilai k yang kecil dapat menyebabkan *overfitting* dan nilai k yang besar dapat menyebabkan *underfitting*. *Overfitting* menyiratkan bahwa model bekerja baik pada saat pelatihan data, tetapi memiliki kinerja yang buruk ketika pengujian model menggunakan data yang baru.

2.4 Evaluasi Kinerja

Performansi atau kinerja model pembelajaran mesin dinilai dari nilai akurasi dan beberapa parameter tambahan seperti presisi, *recall*, dan *F1 score*. Akurasi adalah rasio dari prediksi sebenarnya yang terdiri dari *true positive* (TP) dan *true negative* (TN) terhadap keseluruhan data. Data keseluruhan meliputi penjumlahan dari *true positive* (TP), *false positive* (FP), *true negative* (TN) dan *false negative* (FN). Presisi adalah rasio dari prediksi positif yang benar dibandingkan dengan hasil prediksi positif secara keseluruhan. Sedangkan *recall* adalah perbandingan antara prediksi *true positive* dengan keseluruhan data *true positive*. Jika *classifier* A memiliki *recall* yang lebih tinggi dan *classifier* B memiliki presisi yang lebih tinggi, maka gabungan presisi dan *recall* untuk kedua *classifier* tersebut dapat menjadi satu metrik dengan cara mengambil rata-rata harmoniknya. Parameter ini disebut dengan *F1 score* yang digunakan untuk membandingkan kinerja dua *classifier*. Persamaan (5-8) menjelaskan

Tabel 1. Data Partisipan

Kelompok Usia	Rentang Usia (tahun)	Jenis kelamin (Lk=laki-laki; Pr=Perempuan)	Jumlah (orang)
1	17 – 20	Lk: 1 orang; Pr: 4 orang	5
2	21 – 40	Lk: 4 orang; Pr: 6 orang	10
3	41 – 60	Lk: 5 orang; Pr: 5 orang	10

**Gambar 1. Skematik elektronik sistem klasifikasi usia berjalan**

perhitungan akurasi (*ac*), presisi, *recall*, dan *F1 score*.

$$Akurasi = \frac{TP + TN}{TP + FP + TN + FN} \quad (5)$$

$$Presisi = \frac{TP}{TP + FP} \quad (6)$$

$$Recall = \frac{TP}{TP + FN} \quad (7)$$

$$F1\ score = \frac{2}{\frac{1}{recall} + \frac{1}{precision}} = \frac{2TP}{TP + \frac{1}{2}(FP + FN)} \quad (8)$$

Dalam mengevaluasi atau memilih *classifier* terbaik pada penelitian ini, dilakukan dengan mencari hasil *cross validation* (CV), di mana CV dapat memberikan nilai yang menyatakan kemampuan dalam memperkirakan kinerja model pada data yang tak terlihat dan/atau yang tidak digunakan saat tahap pelatihan [13],[14].

3. Eksperimen

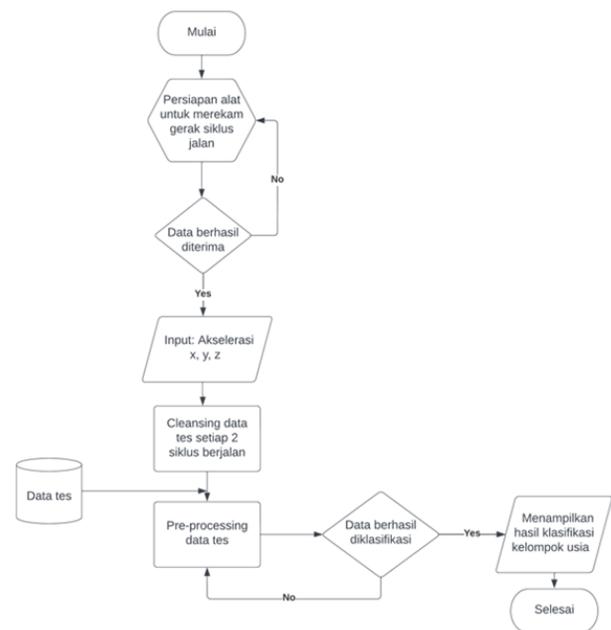
Sub-bagian ini terdiri dari penjelasan tentang partisipan, penjelasan tentang skematik instrumen dan prinsip kerja instrumen yang digunakan, serta skenario eksperimen.

3.1 Partisipan

Sebanyak 25 orang partisipan telah menyetujui untuk melaksanakan pengujian penggunaan sistem prediksi usia berjalan setelah membaca prosedur yang diberikan sesuai dengan kode etik yang berlaku. Tabel 1 menampilkan data demografi singkat dari partisipan dalam proses pengujian sistem, di mana skematik elektronik sistem tersebut dapat dilihat pada Gambar 1.

3.2 Instrumen dan prinsip kerja

Sistem menggunakan mikrokontroler jenis ESP32 dan dua sensor *inertial measurement unit* (IMU) seri MPU9250 yang terhubung dengan multiplexer jenis TCA9548 dalam memudahkan pengiriman data dari kedua sensor secara bersamaan ke mikrokontroler. Data

**Gambar 2. Diagram alir sistem prediksi usia**

selanjutnya dikirim ke PC untuk diproses lebih lanjut dalam penentuan usia seseorang dari gaya berjalannya. Untuk lebih lengkapnya, Gambar 2 menampilkan alur proses akuisisi data dan pengolahan data berjalan untuk memprediksi atau klasifikasi usia seseorang dari gaya berjalannya. Data diambil dan segera disimpan pada penyimpanan internal setiap 2 (dua) siklus berjalan. Segera setelahnya data dihapus dan siap menerima dua data siklus berjalan berikutnya.

3.3 Skenario eksperimen

Dalam memprediksi usia berjalan seseorang dari gaya berjalannya, penelitian ini mencoba menyelidiki apakah usia seseorang dapat ditentukan menggunakan pembacaan sensor inersia pada kedua kaki partisipan selama berjalan dengan hanya menempatkan satu buah



Gambar 3. Ilustrasi penempatan dua sensor MPU9025 pada kedua kaki partisipan

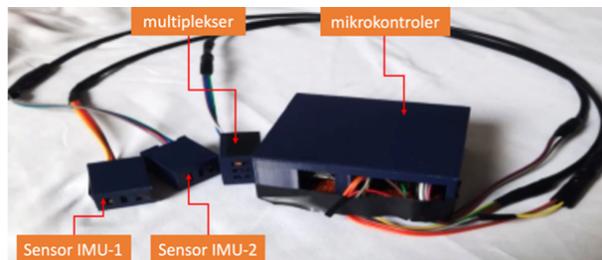
Tabel 2. Hasil akuisisi pembacaan sensor inersia untuk setiap kelompok usia

Kelompok Usia	Jumlah data pembacaan sensor inersia pada kedua kaki
1	10700
2	8500
3	10800

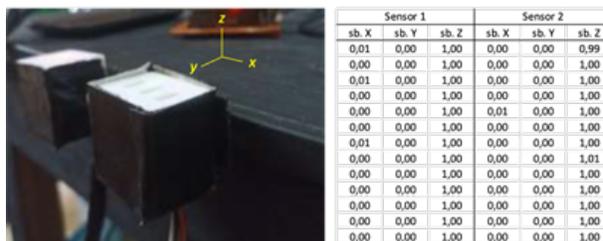
sensor inersia pada masing-masing paha atas dari setiap partisipan, seperti yang terlihat pada Gambar 3. Kemudian, partisipan diminta berjalan selama $\pm 8-10$ menit sambil dilakukan perekaman kedua sensor inersia yang disimpan pada *data logger* lokal dan selanjutnya dipindahkan ke PC untuk proses data berikutnya.

4. Hasil Dan Pembahasan

Hasil perangkat keras dari sistem klasifikasi usia seseorang dari gaya berjalannya ditampilkan pada Gambar 4. Selanjutnya, kedua sensor dipasangkan pada setiap kaki partisipan secara bergantian. Sebelum proses pengukuran dilakukan, proses validasi data sensor MPU9025 perlu dilakukan terlebih dahulu. Hal ini dimaksudkan untuk memastikan akurasi posisi dan akselerasi dari sensor dengan cara menempatkan kedua sensor pada sisi meja untuk merekam data akselerometer pada sumbu x , y , dan z . Dari proses validasi (lihat Gambar 5), dapat dikatakan bahwa kedua sensor memiliki akurasi yang sangat baik, di mana akselerasi sensor di sumbu- x dan sumbu- y adalah nol. Pengambilan data berjalan setiap partisipan dilakukan dalam dua skenario, di mana data yang diperoleh dapat dilihat pada Tabel 2.



Gambar 4. Perangkat keras sistem klasifikasi usia berjalan



Gambar 5. Ilustrasi validasi sensor akselerometer (kiri) dan data pembacaan kedua sensor MPU9025 (kanan)

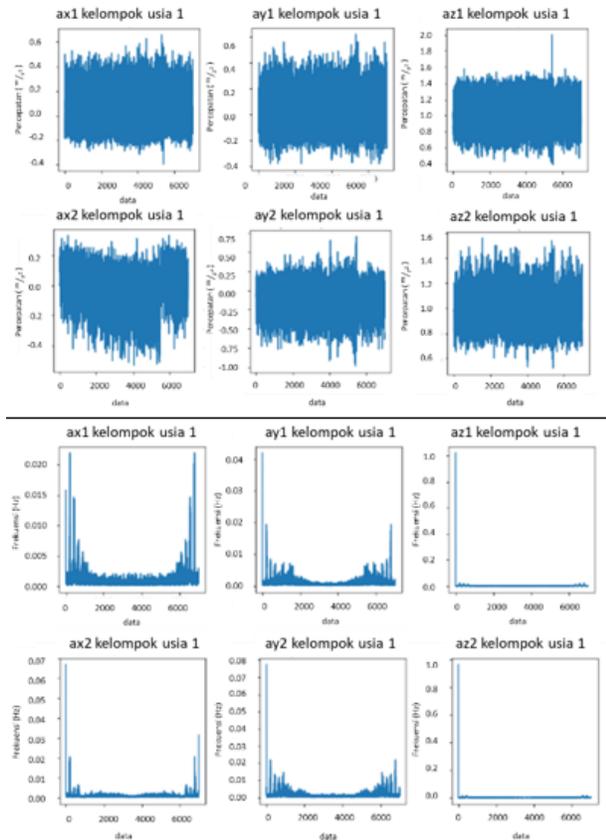
4.1 Hasil pre-processing

Data akselerasi dari kedua kaki setiap partisipan dikumpulkan dan selanjutnya dilakukan *pre-processing*. Fungsi dari *pre-processing* adalah untuk mengubah data sehingga sesuai dengan arsitektur jaringan dan mengurangi dimensi. *Pre-processing* yang dilakukan pada sistem ini dengan menggunakan *Fast Fourier Transform* (FFT). Fungsi dari FFT ini adalah untuk mengurangi dimensi, yaitu mengurangi kompleksitas dari komputasi *Discrete Fourier Transform* (DFT). Proses *pre-processing* ini dilakukan sebelum ekstraksi fitur.

Pada Gambar 6 ditampilkan contoh data akselerasi kedua kaki pada kelompok usia 1 sebelum dan sesudah proses FFT. Sebelum dilakukan proses FFT (Gambar 6a), sinyal masih berupa domain waktu, di mana sinyal terlihat sangat kompleks. Sedangkan setelah dilakukan *pre-processing* dengan menggunakan FFT (Gambar 6b), data menjadi terlihat lebih jelas dan sederhana sehingga lebih mudah untuk dilakukan ekstraksi fitur.

4.2 Pemilihan ekstraksi fitur dan classifier

Sebelum dilakukan proses pelatihan (*training*) dan pengujian (*testing*), *dataset* dipisah dengan proporsi 80:20 (data latih:data uji). Pemilihan komposisi ini juga dilakukan pada penelitian pembelajaran mesin pada umumnya [15]. Sebanyak 10700 data kelompok usia 1 dibagi ke dalam kelompok data latih 80% (8560 data)



Gambar 6. Data akselerasi kedua kaki sebelum (atas) dan sesudah proses FFT (bawah)

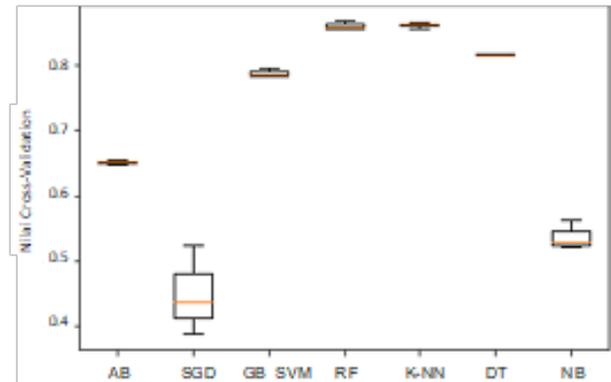
dan data uji sebesar 20% (2140 data). Seperti yang ditampilkan pada Tabel 3, pengelompokan data juga dilakukan untuk kelompok usia 2 dan kelompok usia 3.

Pemilihan jenis ekstraksi fitur dan classifier terbaik dilakukan dengan penilaian evaluasi kinerja. Ekstraksi fitur yang digunakan dalam penelitian ini adalah *mean*, *median*, *max*, *min*, *kurtosis*, dan *skew*. Pada penelitian ini, dipilih ekstraksi fitur tipe *mean* berdasarkan hasil akurasi yang lebih tinggi dibandingkan fitur lainnya, yaitu 85,28% dan fitur mean juga memiliki nilai presisi, *recall*, dan *F1-score* terbaik dari hasil data latih dibandingkan hasil dari fitur lainnya, yaitu secara berturut-turut adalah 0,8530; 0,8528; dan 0,8527, seperti yang ditampilkan pada Tabel 4.

Jenis classifier yang dipilih setelah dilakukan uji perbandingan hasil klasifikasi pada penelitian ini adalah classifier K-NN yang mengungguli jenis classifier lainnya, yaitu AdaBoost, SGD, GB, *random forest*, *decision tree*, dan SVM (lihat Gambar 6) menggunakan metode *cross validation* (CV). Classifier K-NN memiliki nilai CV sebesar 0,85 yang diikuti oleh RF 0,84 serta GB dan SVM sebesar 0,78. Dengan demikian, ekstraksi fitur dan classifier yang digunakan selanjutnya adalah *mean*

Tabel 3. Pembagian data latih dan data uji 80:20 untuk kedua skenario penempatan jumlah sensor pada kaki partisipan

Kelompok Usia	Jumlah data akselerasi kedua kaki	
	Data Latih	Data Uji
1	8560	2140
2	6800	1700
3	8640	2160



Gambar 7. Hasil cross validation (CV) dari enam jenis classifier yang digunakan pada data gaya berjalan

dan K-NN.

4.3 Hasil klasifikasi dengan KNN

Pengujian model sistem prediksi usia berjalan seseorang dilakukan dengan menggunakan data uji sebanyak 6000 data, yang terdiri dari 2140 data kelompok usia 0, 1700 data kelompok usia 1, dan 2160 data kelompok usia 2.

Hasil pengujian model sistem prediksi usia berjalan seseorang yang memperlihatkan performansi classifier pada data ditampilkan dalam *confusion matrix* seperti yang ditampilkan pada Tabel 5. Dari data pada *confusion matrix* dapat diketahui untuk kelompok usia 1 terdapat 154 data yang diprediksi masuk ke kelompok usia 2, di mana hal ini bisa dijadikan sebagai indikasi bahwa 154 data tersebut memiliki permasalahan kesehatan. Sedangkan untuk kelompok 2 dan 3 terdapat data yang terprediksi masuk ke kelompok usia di bawahnya, dengan jumlah 424 data dan 432 data. Hal ini dapat disebut sebagai indikasi bahwa data-data tersebut memiliki kesehatan yang baik karena dikelompokkan ke kelompok usia yang lebih muda dari seharusnya. Selanjutnya, berdasarkan hasil yang ditampilkan pada *confusion matrix*, maka dapat dilihat kinerja model

Tabel 4. Hasil data latih dari 6 (enam) ekstraksi fitur yang berbeda berdasarkan parameter evaluasi kinerja presisi, recall dan F1-score

Ekstraksi fitur	Akurasi (%)	Presisi	Recall	F1-Score
Mean	85,28	0,8530	0,8528	0,8527
Median	55,24	0,5524	0,5530	0,5518
Max	72,45	0,7245	0,7221	0,7225
Min	39,03	0,3903	0,3858	0,3865
Skew	67,95	0,6795	0,6806	0,6796
Kurtosis	60,39	0,6040	0,6034	0,6032

Tabel 5. Confusion matrix data uji

		Prediksi			
		Kelompok Usia			
		1	2	3	
Aktual	Kelompok Usia	1	2006	154	0
		2	424	1276	0
		3	0	432	1728

pembelajaran mesin berdasarkan parameter akurasi, presisi, *recall*, dan *F1-score* sesuai dengan persamaan (5-8) seperti yang ditampilkan pada Tabel 6.

Hasil yang diperoleh dari perekaman sensor inersial berbiaya rendah ini dalam memprediksi usia manusia dari gaya berjalannya adalah sangat menjanjikan untuk dapat dikembangkan dengan lebih baik ke depannya. Akurasi prediksi usia seseorang dari gaya berjalan menggunakan model K-NN dengan ekstraksi fitur *mean* berbasis sensor inersia pada penelitian ini adalah 86,9%. Hasil ini terbilang cukup baik karena hanya menggunakan 25 orang partisipan jika dibandingkan dengan suatu penelitian [16] yang menggunakan populasi data yang sangat besar, yaitu 744 partisipan, dengan akurasi estimasi usianya adalah 85,7% menggunakan pengklasifikasian *random forest*.

5. Kesimpulan

Sistem prediksi usia seseorang telah dapat diprediksi dari gaya berjalan yang diperoleh dari klasifikasi K-NN pada ekstraksi fitur *mean* pembacaan percepatan menggunakan sensor inersia yang dipasang pada kedua kaki partisipan. Fitur *mean* memiliki nilai akurasi terbaik, yaitu 85,28%, dibandingkan fitur lainnya seperti *max*, *skew*, dan lainnya, sedangkan *classifier* K-NN memiliki nilai *cross validation* tertinggi dibandingkan *random forest*, yang memberikan hasil akurasi prediksi usia sebesar 86,9% dengan presisi, *recall*, dan *F1-score* nya sama-sama bernilai 88,0. Untuk penelitian di masa akan datang, sistem akan dikembangkan dengan beberapa variasi, seperti peningkatan akurasi prediksi usia dengan cara

Tabel 6. Performansi model yang diajukan

Parameter	Hasil (%)
Akurasi	86,9
Presisi	88,0
Recall	88,0
F1-Score	88,0

menambah jumlah partisipan, menambah kelompok usia, menggunakan kombinasi multi-fitur, serta komparasi jumlah sensor yang digunakan terhadap akurasi prediksi.

Daftar Pustaka

- [1] W. Tao, T. Liu, R. Zheng, and H. Feng, "Gait analysis using wearable sensors," *Sensors*. 2012.
- [2] M. Hossin and M. N. bin Sulaiman, "A Review On Evaluation Metrics For Data Classification Evaluations," *Int. J. Data Min. & Knowl. Manag. Process*, vol. 5, pp. 1–11, 2015.
- [3] N. Mansouri, M. Aouled Issa, and Y. Ben Jemaa, "Gait-based Human Age Classification Using a Silhouette Model," *IET Biometrics*, vol. 7, Jun. 2017.
- [4] X. Li, Y. Makihara, C. Xu, Y. Yagi, and M. Ren, "Gait-based human age estimation using age group-dependent manifold learning and regression," *Multimed. Tools Appl.*, vol. 77, no. 21, pp. 28333–28354, 2018.
- [5] B. Jin et al., "Walking-age analyzer for healthcare applications," *IEEE J. Biomed. Heal. Informatics*, vol. 18, no. 3, pp. 1034–1042, 2014.
- [6] S. B. Kotsiantis and D. Kanellopoulos, "Data preprocessing for supervised learning," *Int. J. ...*, vol. 1, no. 2, pp. 1–7, 2006.
- [7] Y. Permadi and . Murinto, "Aplikasi Pengolahan Citra Untuk Identifikasi Kematangan Mentimun Berdasarkan Tekstur Kulit Buah Menggunakan Metode Ekstraksi Ciri Statistik," *J. Inform.*, vol. 9, no. 1, pp. 1028–1038, 2015.
- [8] G. I. Gunawan, D. K. Silalahi, H. Mukhtar, D. T. Barus, and D. Rahmawati, "Performance Comparison of Classification Algorithms for Locating the Dominant Heel Pain Using Electromyography Signal," *Lect. Notes Electr. Eng.*, vol. 746 LNEE, pp. 517–529, 2021.
- [9] V. R. Jakkula, "Tutorial on Support Vector Machine (SVM)," 2011.
- [10] A. Satriyo Nugroho, Anto, Handoko, Dwi dan B. Witarto "Analisa Informasi Dimensi Tinggi pada Bioinformatika Memakai Support Vectore Machine,"

- in Proc. of National Conference on Information & Communication Technology (ICT) for Indonesia/e-Indonesia Initiatives-II, 2005, pp. 427-435.
- [11] D. P. Vivencio, E. R. Hruschka, M. Do Carmo Nicoletti, E. B. Dos Santos, and S. D. C. O. Galvão, "Feature-weighted k-nearest neighbor classifier," *Proc. 2007 IEEE Symp. Found. Comput. Intell. FOCI 2007*, no. Foci, pp. 481–485, 2007.
- [12] A. N. Putra, "Studi Algoritma Klasifikasi Sensor Accelerometer dan Gyroscope untuk Pola Activity Daily Life (ADL) pada Dewasa Sehat," *e-Proceeding Eng.*, vol. 5, no. 3, p. 8066, 2018.
- [13] A. Vabalas, E. Gowen, E. Poliakoff, and A. J. Casson, "Machine learning algorithm validation with a limited sample size," *PLoS One*, vol. 14, no. 11, p. e0224365, Nov. 2019.
- [14] B. Ferdinandy et al., "Challenges of machine learning model validation using correlated behaviour data: Evaluation of cross-validation strategies and accuracy measures," *PLoS One*, vol. 15, no. 7, p. e0236092, Jul. 2020.
- [15] A. Gholamy, V. Kreinovich, and O. Kosheleva, "Why 70/30 or 80/20 Relation Between Training and Testing Sets : A Pedagogical Explanation," *Dep. Tech. Reports*, 2018.
- [16] S. I. Gillani, M. A. Azam, and M. Ehatisham-ul-Haq, "Age Estimation and Gender Classification Based on Human Gait Analysis," in *2020 International Conference on Emerging Trends in Smart Technologies (ICETST)*, 2020, pp. 1–6.