

Pengenalan Aktivitas Manusia dengan Seleksi Fitur Analysis of Variance (ANOVA) dan Mutual Information (MI) pada Data Sensor Accelerometer Berbasis Machine Learning

Made Liandana¹, I Made Darma Susila², Yohanes Priyo Atmojo³

Institut Teknologi dan Bisnis STIKOM Bali

e-mail: ¹liandana@stikom-bali.ac.id, ²darma_s@stikom-bali.ac.id, ³yohanes@stikom-bali.ac.id

Diajukan: 23 Desember 2023; Direvisi: 11 Januari 2024; Diterima: 15 Januari 2024

Abstrak

Pengenalan aktivitas manusia telah banyak dikembangkan untuk berbagai keperluan, seperti kesehatan, olahraga, hingga pengawasan lanjut usia. Penggunaan perangkat sensor menjadi salah satu pilihan dalam melakukan pengenalan aktivitas manusia. Sensor accelerometer adalah salah satu perangkat yang umum digunakan dalam pengenalan aktivitas. Data sensor ini memerlukan teknik dan algoritma yang tepat sehingga menghasilkan hasil pengenalan aktivitas yang sesuai. Penggunaan tradisional machine learning menjadi salah satu teknik yang dapat digunakan, teknik ini memerlukan proses ekstraksi fitur, dan seleksi fitur. Teknik seleksi fitur mana dan berapa jumlah fitur yang tepat untuk mendapatkan performa machine learning yang optimal perlu dilakukan investigasi lebih lanjut. Pada penelitian ini, dilakukan evaluasi terhadap kombinasi sejumlah fitur menggunakan algoritma machine learning: Extreme Gradient Boosting (XGB), Gradient Boosting (GBoost), Random Forest (RF), Decision Tree (DT), dan Support Vector Machine (SVM). Dataset publik yang digunakan yaitu FORTH-TRACE. Sensor yang digunakan adalah sensor accelerometer. Fitur yang digunakan meliputi nilai minimum, nilai maksimum, nilai rata-rata, nilai tengah, standar deviasi, dan nilai interkuartil. Sedangkan seleksi fitur yang digunakan adalah Analysis of Variance (ANOVA) dan Mutual Information (MI). Performa machine learning yang paling optimal ketika jumlah fitur 17 sampai dengan 18 fitur dengan akurasi 0,875, sedangkan performa machine learning paling optimal dicapai dengan menggunakan Extreme Gradient Boosting (XGB).

Kata kunci: Accelerometer, MI, ANOVA, Machine learning.

Abstract

The recognition of human activities has been extensively developed for various purposes, such as health, sports, and elderly monitoring. The use of sensor devices has become one of the options for recognizing human activities. An accelerometer sensor is one of the devices commonly used for activity recognition. The data from these sensors require appropriate techniques and algorithms to produce accurate activity recognition results. Traditional machine learning is one technique that can be used, and it requires feature extraction and feature selection processes. Further investigation is needed to determine which feature selection techniques and the number of features are appropriate to achieve optimal machine learning performance. In this study, an evaluation was conducted on the combination of several features using machine learning algorithms: Extreme Gradient Boosting (XGB), Gradient Boosting (GBoost), Random Forest (RF), Decision Tree (DT), and Support Vector Machine (SVM). The public dataset used is FORTH-TRACE. The sensor used is an accelerometer sensor. The features used include minimum, maximum, mean, median, standard deviation, and interquartile range. The feature selection techniques used are Analysis of Variance (ANOVA) and Mutual Information (MI). The most optimal machine learning performance was achieved when using 17 to 18 features, with an accuracy of 0.875, while the most optimal machine learning performance was achieved using Extreme Gradient Boosting (XGB).

Keywords: Accelerometer, MI, ANOVA, Machine learning.

1. Pendahuluan

Pengenalan aktivitas manusia dapat dikenali berdasarkan data sensor dan data gambar atau video [1]–[3][4]–[12]. Data sensor dapat dihasilkan dari sejumlah sensor, seperti: sensor *accelerometer*,

gyroscope, dan magnetometer [13]–[19]. Sedangkan data gambar atau video dihasilkan dari perangkat kamera [20]. Perangkat sensor yang digunakan untuk membaca data dari suatu aktivitas perlu dipasang pada bagian tubuh pengguna, seperti: pinggang, dada, tangan, ataupun kaki [21][22]. Artinya, perangkat tersebut akan menempel dan mengikuti pergerakan dari pengguna. Berbeda halnya dengan penggunaan perangkat kamera, perangkat ini biasanya dipasang pada sudut ruangan atau di tempat yang mampu menjangkau objek yang akan diamati. Penggunaan kamera memiliki keterbatasan dalam mengamati objek yang memiliki mobilitas yang tinggi [8][7], sehingga penggunaan sensor dapat menjadi salah satu pilihan.

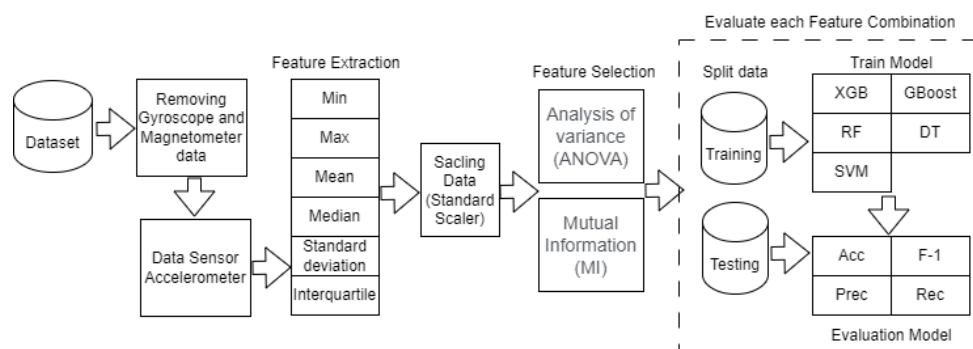
Penggunaan data sensor untuk mengenali aktivitas manusia bukanlah hal yang baru, sejumlah studi telah dilakukan. Data sensor seperti *accelerometer*, *gyroscope*, ataupun *magnetometer* perlu diproses dan diolah dengan teknik atau algoritma yang sesuai [11], [15], [23]. Dalam hal memprediksi suatu aktivitas berdasarkan data sensor, penggunaan *machine learning* telah banyak digunakan [11][23][24]. Penggunaan *machine learning* dapat memberikan hasil prediksi yang lebih akurat dibandingkan dengan cara konvensional. Penggunaan tradisional *machine learning* maupun *deep learning* telah banyak digunakan oleh sejumlah peneliti [11], [15], [23], [25]–[28]. Dalam konteks penggunaan tradisional *machine learning*, penggunaan teknik ekstraksi fitur menjadi bagian yang sangat penting sebelum proses klasifikasi atau prediksi dilakukan [29][30].

Penggunaan ekstraksi fitur dengan pendekatan statistik menjadi salah satu teknik yang cukup populer digunakan, khususnya untuk tradisional *machine learning*. Pendekatan statistik seperti: minimum, maksimum, rata-rata, nilai tengah, dan standar deviasi adalah sejumlah fitur yang dapat digunakan dalam pengenalan aktivitas manusia berbasis sensor [23][31]. Formula matematika untuk menghasilkan nilai statistik tersebut digunakan untuk mengalkulasi data sensor. Sensor seperti *accelerometer* dan *gyroscope* memiliki tiga sumbu sensor, yaitu: sumbu x, y, dan z. Dengan menggunakan sebuah formula matematika, setidaknya akan menghasilkan tiga buah fitur jika data masing-masing sumbu digunakan. Pada sejumlah studi, ketiga sumbu tersebut dapat dikonversi ke dalam sebuah nilai yang dapat disebut dengan istilah total akselerasi atau magnitude [32][33].

Penggunaan tradisional *machine learning*, selain memerlukan proses ekstraksi fitur, diperlukan juga tahapan seleksi fitur. Seleksi fitur dapat membantu mereduksi jumlah fitur sehingga dapat mengurangi proses komputasi [25]. Penelitian sebelumnya menunjukkan, penggunaan seleksi fitur dapat meningkatkan performa *machine learning* dalam mengklasifikasi suatu aktivitas [34] [35]. Pada penelitian ini, seleksi fitur yang digunakan adalah Analysis of Variance (ANOVA) dan Mutual Information (MI). Penelitian sebelumnya, menunjukkan seleksi fitur ini dapat memberikan dampak yang signifikan pada kemampuan *machine learning* untuk mengklasifikasi data aktivitas. Namun, bagaimana pengaruh kedua seleksi fitur ini jika dievaluasi berdasarkan setiap kombinasi fitur yang dihasilkan menjadi hal yang perlu diinvestigasi lebih lanjut. Sehingga pada penelitian ini, dilakukan investigasi lebih lanjut terhadap kedua seleksi fitur ini dengan cara mengevaluasi setiap kombinasi fitur yang dihasilkan dari kedua seleksi fitur ini. Tujuannya adalah untuk mengetahui kombinasi fitur mana yang dapat memberikan dampak yang signifikan terhadap performa *machine learning*, khususnya kombinasi fitur yang dihasilkan dengan menggunakan seleksi fitur Analysis of Variance (ANOVA) dan Mutual Information (MI).

2. Metode Penelitian

Data yang bersumber dari dataset publik sering kali terdiri dari data sensor *accelerometer*, *gyroscope*, dan *magnetometer*, pada penelitian yang dilakukan hanya data sensor yang bersumber dari sensor *accelerometer* yang akan diproses. Seperti pada Gambar 1, model yang diusulkan terdiri dari sejumlah tahapan atau proses.



Gambar 1. Model yang diusulkan.

Proses yang pertama dilakukan adalah memilih data sensor *accelerometer* dari sekumpulan data sensor yang ada pada dataset. Selanjutnya proses ekstraksi fitur dilakukan, sehingga dihasilkan fitur-fitur penting, yaitu: nilai minimum, nilai maksimum, nilai rata-rata, nilai tengah, standar deviasi, dan nilai interkuartil. Dengan menggunakan sensor *accelerometer* 3 sumbu, maka akan dihasilkan 3 kali lipat jumlah fitur, yaitu sebanyak 18 fitur. Untuk menghindari nilai dominasi dari setiap fitur, maka diperlukan tahapan *scaling*, pada kasus ini menggunakan *standard scaler*. Untuk memilih fitur mana yang memberikan dampak signifikan terhadap performa *machine learning* dalam mengklasifikasi aktivitas, peranan fitur seleksi sangat diperlukan. Metode seleksi fitur yang digunakan adalah Analysis of Variance (ANOVA) dan Mutual Information (MI). Kedua metode ini akan diuji menggunakan algoritma *machine learning*, yaitu: Extreme Gradient Boosting (XGB), Gradient Boosting (GBoost), Random Forest (RF), Decision Tree (DT), dan Support Vector Machine (SVM). Performa dari *machine learning* diukur menggunakan nilai *accuracy*, *precision*, *recall*, dan *F-1 score*. Studi yang dilakukan melakukan proses evaluasi pada setiap jumlah kombinasi fitur, mulai dari kombinasi 1 fitur hingga jumlah maksimal fitur.

2.1. Ekstraksi Fitur

Penggunaan tradisional *machine learning* memerlukan tahapan ekstraksi fitur, tahapan ini menjadi bagian yang sangat penting karena akan menghasilkan ciri dari suatu data yang diklasifikasi. Pada penelitian ini, ekstraksi fitur dilakukan menggunakan perhitungan statistik. Data sensor *accelerometer* diekstraksi menjadi nilai minimum (*min*), maksimum (*max*), rata-rata (Mean), nilai tengah (*Median*), standar deviasi (*Std*), dan interkuartil (*Iqr*), secara kronologis menggunakan Persamaan (1), Persamaan (2), Persamaan (3), Persamaan (4), Persamaan (5), dan Persamaan (6). *Acc* merupakan nilai akselerasi, untuk sumbu x, y, dan z. \overline{Acc} merupakan nilai rata-rata, Med_{Acc} merupakan nilai tengah, Std_{Acc} merupakan nilai standar deviasi. Untuk Q_1 merupakan kuartil ke-1, dan Q_3 merupakan kuartil ke-3. Sedangkan *i* adalah data ke-*i* dari sejumlah data serial yang akan dievaluasi.

$$Min_{Acc} = \min(Acc_i) \tag{1}$$

$$Max = \max(Acc_i) \tag{2}$$

$$\overline{Acc} = \frac{1}{n} \sum_{i=1}^n (Acc_i) \tag{3}$$

$$Med_{Acc} = \begin{cases} Acc_{\lfloor \frac{n+1}{2} \rfloor}, & \text{if } n \text{ odd} \\ \frac{1}{2} (Acc_{\lfloor \frac{n}{2} \rfloor} + Acc_{\lfloor \frac{n+1}{2} \rfloor}), & \text{if } n \text{ even} \end{cases} \tag{4}$$

$$Std_{Acc} = \sqrt{\frac{1}{n} \sum_{i=1}^n (Acc_i - \overline{Acc})^2} \tag{5}$$

$$Iqr_{Acc} = Q_3 - Q_1 \tag{6}$$

2.2. Seleksi Fitur ANOVA (Analysis of Variance)

Seleksi fitur menggunakan algoritma ANOVA (Analysis of Variance), perhitungannya menggunakan Persamaan (7). Variabel *i* bernilai 1 sampai *k*, *k* merupakan banyaknya kelompok yang dikelompokkan dari semua sampel data. Sedangkan \overline{Acc}_i adalah rata-rata dari setiap kelompok, gd_{ij} merupakan *sample* data dari setiap kelompok, panjang setiap kelompok adalah n_i . \overline{Acc} adalah rata-rata dari semua *sample* dan Gd_m merupakan semua *sample* data yang jumlah sebanyak *K*.

$$skor = \frac{K - k}{k - 1} \times \frac{\sum_{i=1}^k n_i (\overline{Acc}_i - \overline{Acc})^2}{\sum_{i=1}^k \sum_{j=1}^{n_i} (gd_{ij} - \overline{Acc}_i)^2} \quad (7)$$

$$\overline{Acc}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} gd_{ij} \quad (8)$$

$$\overline{Acc} = \frac{1}{K} \sum_{m=1}^K Gd_m \quad (9)$$

2.3. Mutual Information (MI)

Pendekatan lain yang digunakan untuk menyeleksi fitur adalah Mutual Information (MI). Formula yang digunakan untuk menghitung Mutual Information (MI) yaitu Persamaan (10). Dengan $p(x, y)$ adalah *joint probability*. Sedangkan $p(x)$ dan $p(y)$ adalah *marginal probability*. Dalam persamaan (10) diilustrasikan menggunakan dua buah fitur.

$$MI = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log \left[\frac{p(x, y)}{p(x)p(y)} \right] \quad (10)$$

3. Hasil dan Pembahasan

3.1. Deskripsi Dataset

Pada studi ini, dataset yang digunakan adalah dataset FORTH-TRACE [36]. Dataset ini diambil dari 15 orang menggunakan sensor *accelerometer*, *gyroscope*, dan *magnetometer* dengan rata-rata pencuplikan data sebesar 51.2 Hz. Setiap orang melakukan 16 aktivitas, meliputi: tujuh aktivitas dasar/*basic activity* dan sembilan aktivitas untuk sikap badan transisi/*postural transision activity*. Untuk pertimbangan dimensi data, data yang digunakan hanya data sensor *accelerometer*. Sedangkan aktivitas yang digunakan dalam eksperimen ini adalah: berdiri, duduk, berjalan, menaiki tangga, berdiri diikuti duduk, dan duduk diikuti berdiri.

3.2. Penentuan Jumlah Fitur Berdasarkan Menggunakan ANOVA (Analysis of Variance) dan Mutual Information (MI)

Tabel 1 adalah fitur-fitur yang sudah terbentuk berdasarkan data sensor *accelerometer*. Fitur yang dihasilkan dihitung menggunakan Persamaan (1) sampai dengan Persamaan (6). Dengan jumlah sumbu sensor sebanyak 3 akan dihasilkan sebanyak 18 fitur seperti ditunjukkan pada Tabel 1. Untuk mengetahui fitur mana yang memberikan dampak signifikan terhadap performa *machine learning*, peran dari algoritma seleksi fitur ANOVA (Analysis of Variance) dan Mutual Information (MI) diperlukan untuk menyeleksi fitur. Fitur yang terdiri dari 18 fitur diberikan peringkat dengan menggunakan Analysis of Variance (ANOVA) dan Mutual Information (MI). Pembentukan kombinasi fitur, diawali dari fitur dengan peringkat tertinggi. Fitur dengan peringkat tertinggi dipilih menjadi kombinasi fitur pertama, kemudian untuk kombinasi fitur kedua terdiri dari fitur dengan nilai peringkat tertinggi dan fitur dengan nilai peringkat kedua. Hal yang serupa juga berlaku untuk pembentukan kombinasi fitur yang lain, sampai terbentuk 18 kombinasi fitur. Selanjutnya, setiap kombinasi fitur akan dievaluasi menggunakan algoritma *machine learning*.

Tabel 1. Nama-nama Fitur.

No	Fitur	Keterangan	No	Fitur	Keterangan
1	<i>x_min</i>	Nilai minimum untuk sumbu <i>x</i>	10	<i>y_med</i>	Nilai tengah untuk sumbu <i>y</i>
2	<i>x_max</i>	Nilai maksimum untuk sumbu <i>x</i>	11	<i>y_std</i>	Standar deviasi untuk sumbu <i>y</i>
3	<i>x_mean</i>	Rata-rata untuk sumbu <i>x</i>	12	<i>y_iqr</i>	Interkuartil untuk sumbu <i>y</i>
4	<i>x_med</i>	Nilai tengah untuk sumbu <i>x</i>	13	<i>z_min</i>	Nilai minimum untuk sumbu <i>z</i>
5	<i>x_std</i>	Standar deviasi untuk sumbu <i>x</i>	14	<i>z_max</i>	Nilai maksimum untuk sumbu <i>z</i>
6	<i>x_iqr</i>	Interkuartil untuk sumbu <i>x</i>	15	<i>z_mean</i>	Rata-rata untuk sumbu <i>z</i>
7	<i>y_min</i>	Nilai minimum untuk sumbu <i>y</i>	16	<i>z_med</i>	Nilai tengah untuk sumbu <i>z</i>
8	<i>y_max</i>	Nilai maksimum untuk sumbu <i>y</i>	17	<i>z_std</i>	Standar deviasi untuk sumbu <i>z</i>
9	<i>y_mean</i>	Rata-rata untuk sumbu <i>y</i>	18	<i>z_iqr</i>	Interkuartil untuk sumbu <i>z</i>

3.3. Performa Machine Learning

Tabel 2 hingga Tabel 6 merupakan performa dari *machine learning* Extreme Gradient Boosting (XGB), Gradient Boosting (GBoost), Random Forest (RF), Decision Tree (DT), dan Support Vector Machine (SVM) untuk penggunaan seleksi fitur ANOVA. Sedangkan Tabel 7 sampai dengan Tabel 11 adalah performa *machine learning* yang sama, namun dengan menggunakan seleksi fitur Mutual Information. Dari tabel tersebut menunjukkan, performa optimal dari sebagian besar algoritma *machine learning* yang digunakan dicapai pada kombinasi fitur sebanyak 17 dan 18 fitur, baik untuk penggunaan seleksi fitur ANOVA maupun untuk seleksi fitur Mutual Information (MI). Hasil eksperimen tersebut menunjukkan, penggunaan fitur seleksi Mutual Information (MI) unggul dibandingkan dengan ANOVA terutama untuk jumlah fitur yang lebih sedikit, hal ini dapat dilihat dari penggunaan 1 atau 5 fitur. Akan tetapi, untuk jumlah fitur yang lebih banyak, misalnya penggunaan 17 sampai 18 fitur, penggunaan kedua seleksi fitur ini tidak memiliki perbedaan yang signifikan.

Tabel 2. Performa Extreme Gradient Boosting (XGB) dengan Analysis of Variance (ANOVA).

Jumlah Fitur	Accuracy	Precision	Recall	F1-Score
1	0,627	0,313	0,380	0,338
2	0,638	0,402	0,385	0,354
3	0,656	0,358	0,393	0,360
4	0,675	0,573	0,424	0,410
5	0,772	0,606	0,558	0,562
6	0,808	0,648	0,584	0,603
7	0,822	0,670	0,594	0,616
8	0,833	0,692	0,612	0,635
9	0,842	0,670	0,623	0,634
10	0,845	0,680	0,622	0,636
11	0,845	0,661	0,629	0,632
12	0,853	0,699	0,642	0,656
13	0,866	0,707	0,651	0,669
14	0,862	0,717	0,651	0,673
15	0,867	0,712	0,655	0,674
16	0,873	0,709	0,640	0,663
17	0,875	0,696	0,638	0,658
18	0,875	0,719	0,650	0,673

Tabel 3. Performa Gradient Boosting (GBoost) dengan Analysis of Variance (ANOVA).

<i>Jumlah Fitur</i>	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
1	0,623	0,365	0,382	0,341
2	0,629	0,329	0,379	0,344
3	0,648	0,387	0,397	0,369
4	0,654	0,541	0,410	0,389
5	0,768	0,627	0,539	0,563
6	0,788	0,642	0,569	0,588
7	0,804	0,648	0,566	0,588
8	0,813	0,676	0,586	0,609
9	0,818	0,671	0,590	0,612
10	0,821	0,678	0,593	0,615
11	0,829	0,677	0,616	0,631
12	0,832	0,715	0,624	0,648
13	0,831	0,686	0,627	0,647
14	0,837	0,719	0,622	0,650
15	0,835	0,713	0,619	0,647
16	0,840	0,710	0,624	0,649
17	0,840	0,693	0,613	0,637
18	0,840	0,701	0,620	0,645

Tabel 4. Performa Random Forest dengan Analysis of Variance (ANOVA).

<i>Jumlah Fitur</i>	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
1	0,535	0,347	0,343	0,342
2	0,604	0,438	0,368	0,348
3	0,646	0,496	0,393	0,366
4	0,679	0,630	0,429	0,425
5	0,781	0,656	0,545	0,569
6	0,799	0,679	0,544	0,566
7	0,816	0,644	0,565	0,586
8	0,821	0,682	0,567	0,590
9	0,835	0,723	0,588	0,608
10	0,839	0,641	0,587	0,604
11	0,842	0,673	0,594	0,616
12	0,848	0,672	0,601	0,623
13	0,852	0,711	0,602	0,626
14	0,849	0,643	0,604	0,616
15	0,853	0,676	0,601	0,615
16	0,859	0,681	0,606	0,629
17	0,859	0,648	0,603	0,617
18	0,865	0,698	0,611	0,635

Tabel 5. Performa Decision Tree dengan Analysis of Variance (ANOVA).

<i>Jumlah Fitur</i>	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
1	0,534	0,349	0,347	0,346
2	0,555	0,371	0,374	0,364
3	0,618	0,445	0,410	0,404
4	0,624	0,459	0,425	0,415
5	0,669	0,507	0,503	0,479
6	0,699	0,530	0,532	0,508
7	0,759	0,587	0,580	0,578
8	0,722	0,539	0,557	0,538
9	0,748	0,556	0,575	0,558
10	0,730	0,549	0,553	0,542
11	0,761	0,580	0,602	0,580
12	0,745	0,564	0,580	0,563
13	0,745	0,568	0,583	0,561
14	0,767	0,572	0,586	0,568
15	0,764	0,571	0,587	0,568
16	0,763	0,573	0,592	0,569
17	0,764	0,574	0,590	0,570
18	0,768	0,580	0,596	0,571

Tabel 6. Performa Support Vector Machine (SVM) dengan Analysis of Variance (ANOVA).

<i>Jumlah Fitur</i>	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
1	0,630	0,313	0,377	0,338
2	0,641	0,328	0,370	0,332
3	0,656	0,341	0,383	0,346
4	0,663	0,449	0,409	0,388
5	0,777	0,674	0,532	0,561
6	0,795	0,707	0,546	0,576
7	0,809	0,727	0,564	0,596
8	0,815	0,731	0,575	0,609
9	0,825	0,741	0,586	0,620
10	0,828	0,743	0,588	0,622
11	0,835	0,748	0,595	0,628
12	0,835	0,747	0,595	0,628
13	0,840	0,759	0,599	0,631
14	0,839	0,753	0,598	0,629
15	0,840	0,755	0,601	0,635
16	0,844	0,710	0,601	0,630
17	0,848	0,711	0,604	0,634
18	0,850	0,761	0,610	0,643

Tabel 7. Performa Extreme Gradient Boosting dengan Mutual Information (MI).

<i>Jumlah Fitur</i>	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
1	0,724	0,488	0,469	0,472
2	0,747	0,631	0,515	0,542
3	0,768	0,638	0,536	0,561
4	0,794	0,656	0,564	0,590
5	0,801	0,642	0,569	0,591
6	0,817	0,691	0,595	0,621
7	0,828	0,686	0,596	0,621
8	0,832	0,670	0,592	0,614
9	0,837	0,674	0,616	0,633
10	0,845	0,680	0,622	0,636
11	0,860	0,702	0,636	0,659
12	0,863	0,715	0,647	0,671
13	0,866	0,698	0,638	0,659
14	0,863	0,697	0,637	0,658
15	0,866	0,711	0,644	0,668
16	0,873	0,709	0,640	0,663
17	0,876	0,724	0,656	0,679
18	0,875	0,719	0,650	0,673

Tabel 8. Performa Gradient Boosting dengan Mutual Information (MI).

<i>Jumlah Fitur</i>	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
1	0,719	0,539	0,483	0,493
2	0,749	0,596	0,515	0,536
3	0,765	0,625	0,547	0,571
4	0,775	0,623	0,554	0,574
5	0,786	0,582	0,539	0,551
6	0,794	0,640	0,563	0,582
7	0,805	0,647	0,571	0,588
8	0,810	0,680	0,581	0,607
9	0,819	0,718	0,592	0,623
10	0,821	0,678	0,593	0,615
11	0,824	0,688	0,599	0,624
12	0,827	0,683	0,597	0,622
13	0,834	0,743	0,611	0,645
14	0,833	0,673	0,597	0,620
15	0,839	0,733	0,632	0,656
16	0,841	0,714	0,624	0,649
17	0,841	0,700	0,621	0,646
18	0,840	0,704	0,620	0,646

Tabel 9. Performa Random Forest dengan Mutual Information (MI).

<i>Jumlah Fitur</i>	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
1	0,637	0,455	0,457	0,453
2	0,718	0,592	0,500	0,523
3	0,747	0,666	0,519	0,547
4	0,781	0,672	0,542	0,568
5	0,801	0,685	0,556	0,579
6	0,812	0,700	0,564	0,591
7	0,822	0,672	0,570	0,595
8	0,823	0,659	0,564	0,586
9	0,833	0,724	0,593	0,622
10	0,839	0,641	0,587	0,604
11	0,845	0,697	0,596	0,619
12	0,850	0,698	0,600	0,626
13	0,851	0,750	0,597	0,621
14	0,851	0,647	0,600	0,615
15	0,855	0,703	0,601	0,623
16	0,859	0,681	0,606	0,629
17	0,861	0,668	0,608	0,627
18	0,865	0,698	0,611	0,635

Tabel 10. Performa Decision Tree dengan Mutual Information (MI).

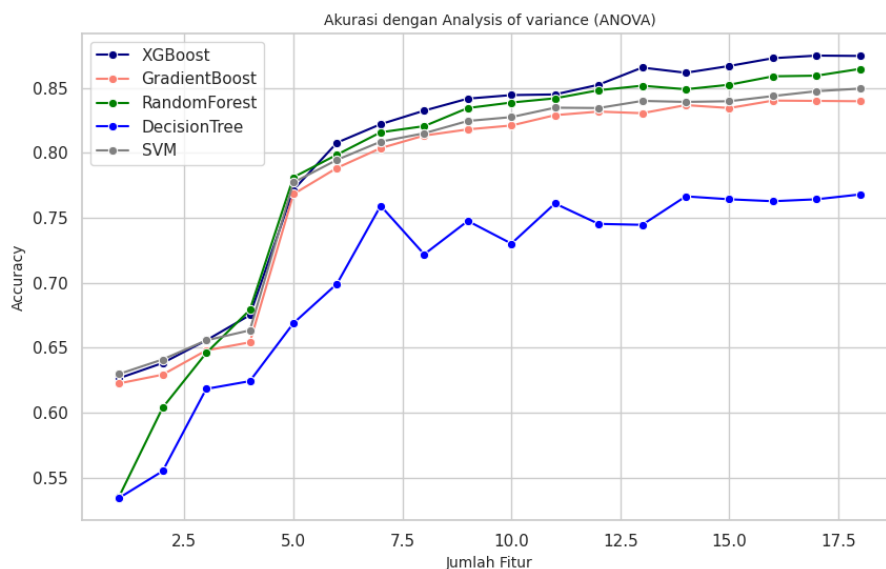
<i>Jumlah Fitur</i>	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
1	0,610	0,439	0,452	0,440
2	0,664	0,492	0,483	0,484
3	0,685	0,513	0,504	0,504
4	0,692	0,514	0,514	0,510
5	0,722	0,542	0,524	0,529
6	0,756	0,572	0,562	0,563
7	0,761	0,584	0,591	0,580
8	0,739	0,560	0,564	0,543
9	0,744	0,566	0,582	0,564
10	0,730	0,549	0,553	0,542
11	0,755	0,568	0,575	0,564
12	0,763	0,585	0,605	0,585
13	0,757	0,568	0,578	0,563
14	0,758	0,567	0,580	0,559
15	0,775	0,579	0,594	0,573
16	0,763	0,573	0,592	0,569
17	0,768	0,581	0,589	0,573
18	0,768	0,580	0,596	0,571

Gambar 2 adalah performa dari sisi akurasi (*accuracy*) untuk 5 model *machine learning*, ditinjau dari jumlah fitur yang digunakan, khususnya untuk penerapan seleksi fitur ANOVA. Dari diagram tersebut

menunjukkan performa Extreme Gradient Boosting (XGBoost) paling optimal dibandingkan dengan algoritma *machine learning* lain yang digunakan. Jumlah fitur yang paling optimal untuk *machine learning* XGBoost adalah sebanyak 17 dan 18 fitur.

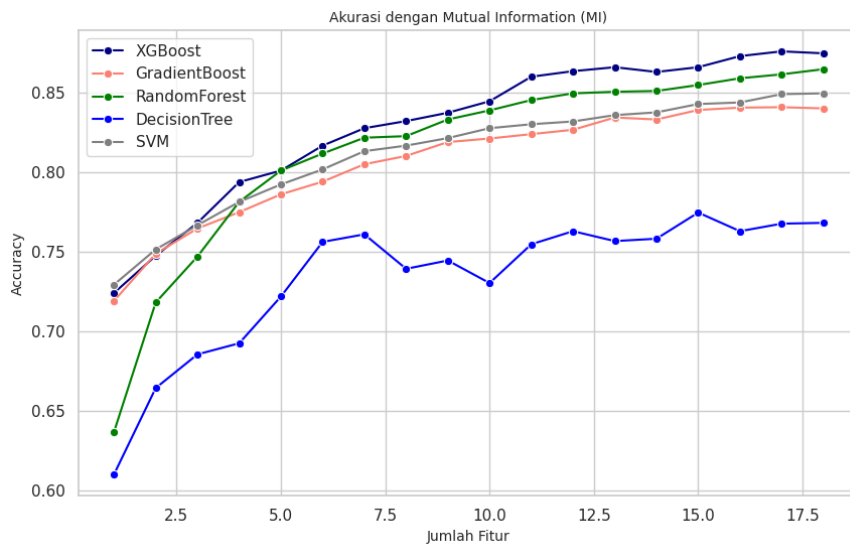
Tabel 11. Performa Support Vector *Machine* (SVM) dengan Mutual Information (MI).

Jumlah Fitur	Accuracy	Precision	Recall	F1-Score
1	0,729	0,493	0,466	0,471
2	0,751	0,609	0,489	0,506
3	0,767	0,628	0,510	0,528
4	0,781	0,661	0,533	0,559
5	0,792	0,671	0,540	0,566
6	0,802	0,680	0,561	0,590
7	0,813	0,690	0,570	0,599
8	0,817	0,700	0,573	0,603
9	0,821	0,731	0,580	0,614
10	0,828	0,743	0,588	0,622
11	0,830	0,738	0,590	0,624
12	0,832	0,749	0,595	0,629
13	0,836	0,705	0,594	0,624
14	0,838	0,704	0,592	0,621
15	0,843	0,707	0,597	0,626
16	0,844	0,710	0,601	0,630
17	0,849	0,761	0,609	0,643
18	0,850	0,761	0,610	0,643



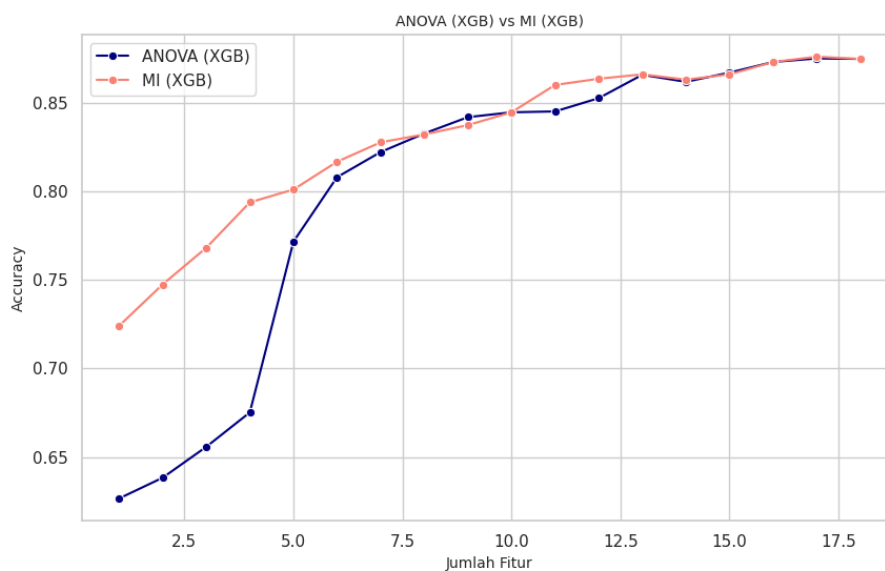
Gambar 2. Perbandingan Performa Machine Learning Menggunakan Seleksi Fitur ANOVA.

Sedangkan Gambar 3 merupakan akurasi (*accuracy*) 5 model *machine learning* khususnya untuk penggunaan seleksi fitur Mutual Information (MI). Gambar tersebut menunjukkan performa *machine learning* Extreme Gradient Boosting (XGBoost) memiliki performa yang paling baik di antara 4 algoritma yang digunakan.



Gambar 3. Perbandingan Performa *Machine Learning* Menggunakan Seleksi Fitur Mutual Information (MI).

Gambar 4 menunjukkan perbedaan akurasi (*accuracy*) khusus untuk algoritma Extreme Gradient Boosting (XGB) untuk penggunaan seleksi fitur ANOVA dan Mutual Information (MI). Untuk penggunaan jumlah fitur yang sedikit (1 hingga 6 fitur), Seleksi fitur Mutual Information (MI) menghasilkan nilai akurasi yang lebih tinggi dibandingkan penggunaan seleksi fitur ANOVA, namun setelah jumlah fitur ditingkatkan penggunaan keduanya dapat menghasilkan nilai akurasi yang hampir sama.



Gambar 4. Perbandingan Extreme Gradient Boosting Menggunakan ANOVA dan Mutual Information (MI).

4. Kesimpulan

Pada penelitian ini, aktivitas yang diklasifikasi menggunakan *machine learning* yaitu: berdiri, duduk, berjalan, menaiki tangga, berdiri diikuti duduk, dan duduk diikuti berdiri, data aktivitas tersebut berdasarkan data dari sensor *accelerometer*. Dari hasil performa *machine learning*, seleksi fitur Mutual

Information (MI) dapat memberikan pengaruh yang lebih signifikan pada kombinasi fitur yang lebih sedikit dibandingkan dengan seleksi fitur Analysis of variance (ANOVA), hal ini terjadi pada jumlah fitur mulai dari 1 fitur hingga 5 fitur. Akan tetapi, memberikan dampak performa *machine learning* yang hampir sama, ketika kombinasi fitur yang digunakan lebih banyak, yaitu pada jumlah fitur 17 sampai dengan 18 fitur.

Selain itu, dari 5 algoritma *machine learning* yang digunakan, algoritma Extreme Gradient Boosting (XGB) memiliki performa yang lebih baik dibandingkan dengan algoritma *machine learning* lain yang digunakan, seperti: Gradient Boosting (Gboost), Random Forest (RF), Decision Tree (DT), dan Support Vector *Machine* (SVM).

Daftar Pustaka

- [1] G. A. S. Surek, L. O. Seman, S. F. Stefenon, V. C. Mariani, and L. dos S. Coelho, "Video-Based Human Activity Recognition Using Deep *Learning* Approaches," *Sensors* 2023, Vol. 23, Page 6384, vol. 23, no. 14, p. 6384, Jul. 2023.
- [2] M. Tabish, Z. ur R. Tanooli, and M. Shaheen, "Activity recognition framework in sports videos," *Multimed Tools Appl*, pp. 1–23, Feb. 2021.
- [3] A. Hussain, T. Hussain, W. Ullah, and S. W. Baik, "Vision Transformer and Deep Sequence *Learning* for Human Activity Recognition in Surveillance Videos," *Comput Intell Neurosci*, vol. 2022, pp. 1–10, Apr. 2022.
- [4] A. M. A. Baraka and M. H. Mohd Noor, "Similarity Segmentation Approach for Sensor-Based Activity Recognition," *IEEE Sens J*, vol. 23, no. 17, pp. 19704–19716, Sep. 2023.
- [5] M. Munoz-Organero, I. D. Luptáková, M. Kubovčík, and J. Pospíchal, "Wearable Sensor-Based Human Activity Recognition with Transformer Model," *Sensors* 2022, Vol. 22, Page 1911, vol. 22, no. 5, p. 1911, Mar. 2022.
- [6] F. Rustam, A. A. Reshi, I. Ashraf, A. Mehmood, S. Ullah, D. M. Khan, and G. S. Choi, "Sensor-Based Human Activity Recognition Using Deep Stacked Multilayered Perceptron Model," *IEEE Access*, vol. 8, pp. 218898–218910, 2020.
- [7] F. Serpush, M. B. Menhaj, B. Masoumi, and B. Karasfi, "Wearable Sensor-Based Human Activity Recognition in the Smart Healthcare System," *Comput Intell Neurosci*, vol. 2022, 2022.
- [8] H. Wang, J. Zhao, J. Li, L. Tian, P. Tu, T. Cao, Y. An, K. Wang, and S. Li, "Wearable Sensor-Based Human Activity Recognition Using Hybrid Deep *Learning* Techniques," *Security and Communication Networks*, vol. 2020, 2020.
- [9] F. Serpush, M. B. Menhaj, B. Masoumi, and B. Karasfi, "Wearable Sensor-Based Human Activity Recognition in the Smart Healthcare System," *Comput Intell Neurosci*, vol. 2022, 2022.
- [10] H. Wang, J. Zhao, J. Li, L. Tian, P. Tu, T. Cao, Y. An, K. Wang, and S. Li, "Wearable Sensor-Based Human Activity Recognition Using Hybrid Deep *Learning* Techniques," *Security and Communication Networks*, vol. 2020, 2020.
- [11] A. B. Mesanza, S. Lucas, A. Zubizarreta, I. Cabanes, E. Portillo, and A. Rodriguez-Larrad, "A *Machine Learning* Approach to Perform Physical Activity Classification Using a Sensorized Crutch Tip," *IEEE Access*, vol. 8, pp. 210023–210034, 2020.
- [12] R. A. Voicu, C. Dobre, L. Bajenaru, and R. I. Ciobanu, "Human Physical Activity Recognition Using Smartphone Sensors," *Sensors (Basel)*, vol. 19, no. 3, Feb. 2019.
- [13] J. Zhu, W. Wang, S. Huang, and W. Ding, "An Improved Calibration Technique for MEMS *Accelerometer*-Based Inclinometers," *Sensors* 2020, Vol. 20, Page 452, vol. 20, no. 2, p. 452, Jan. 2020.
- [14] J. Lu and K. Y. Tong, "Robust Single *Accelerometer*-Based Activity Recognition Using Modified Recurrence Plot," *IEEE Sens J*, vol. 19, no. 15, pp. 6317–6324, Aug. 2019.
- [15] T. Althobaiti, S. Katsigiannis, and N. Ramzan, "Triaxial *Accelerometer*-Based Falls and Activities of Daily Life Detection Using *Machine Learning*," *Sensors* 2020, Vol. 20, Page 3777, vol. 20, no. 13, p. 3777, Jul. 2020.
- [16] E. Fridriksdottir and A. G. Bonomi, "*Accelerometer*-Based Human Activity Recognition for Patient Monitoring Using a Deep Neural Network," *Sensors* 2020, Vol. 20, Page 6424, vol. 20, no. 22, p. 6424, Nov. 2020.
- [17] E. Fortune, V. Lugade, S. Amin, and K. R. Kaufman, "Step detection using multi- versus single tri-axial *accelerometer*-based systems," *Physiol Meas*, vol. 36, no. 12, pp. 2519–2535, Nov. 2015.
- [18] E. Al-Mahadeen, M. Alghamdi, A. S. Tarawneh, M. A. Alrowaily, M. Alrashidi, I. S. Alkhazi, A. Mbaidin, A. A. Alkasasbeh, M. A. Abbadi, and A. B. Hassanat, "Smartphone User Identification/Authentication Using *Accelerometer* and Gyroscope Data," *Sustainability* 2023, Vol. 15, Page 10456, vol. 15, no. 13, p. 10456, Jul. 2023.

-
- [19] W. M. Niu, F. Li-Qing, Z. Y. Qi, and D. Q. Guo, "Small Displacement Measuring System Based on MEMS Accelerometer," *Math Probl Eng*, vol. 2019, 2019.
- [20] F. Taffoni, D. Rivera, A. La Camera, A. Nicolò, J. R. Velasco, and C. Massaroni, "A Wearable System for Real-Time Continuous Monitoring of Physical Activity," *J Healthc Eng*, vol. 2018, pp. 1–16, Mar. 2018.
- [21] B. McDevitt, J. Connolly, D. Duddy, R. Doherty, and J. Condell, "Preliminary Investigations of the Validity and Interinstrument Reliability for Classification of Accelerometer Physical Activity Cut-Points Against Indirect Calorimetry in Healthy Adults," *2022 33rd Irish Signals and Systems Conference, ISSC 2022*, 2022.
- [22] G. I. Mielke, M. de Almeida Mendes, U. Ekelund, A. V. Rowlands, F. F. Reichert, and I. Crochemore-Silva, "Absolute intensity thresholds for tri-axial wrist and waist accelerometer-measured movement behaviors in adults," *Scand J Med Sci Sports*, vol. 33, no. 9, pp. 1752–1764, Sep. 2023.
- [23] B. Vidya and S. P., "Wearable multi-sensor data fusion approach for human activity recognition using machine learning algorithms," *Sens Actuators A Phys*, vol. 341, p. 113557, Jul. 2022.
- [24] W. Xiao and Y. Lu, "Daily Human Physical Activity Recognition Based on Kernel Discriminant Analysis and Extreme Learning Machine," *Math Probl Eng*, vol. 2015, p. 8, 2015.
- [25] B. Nguyen, Y. Coelho, T. Bastos, and S. Krishnan, "Trends in human activity recognition with focus on machine learning and power requirements," *Machine Learning with Applications*, vol. 5, p. 100072, Sep. 2021.
- [26] A. K. Alhazmi, M. A. Alanazi, C. Liu, and V. P. Chodavarapu, "Machine Learning Enabled Fall Detection with Compact Millimeter Wave System," *Proceedings of the IEEE National Aerospace Electronics Conference, NAECON*, vol. 2021-August, pp. 217–222, 2021.
- [27] S. Mohsen, A. Elkaseer, and S. G. Scholz, "Human Activity Recognition Using K-Nearest Neighbor Machine Learning Algorithm," *Smart Innovation, Systems and Technologies*, vol. 262 SIST, pp. 304–313, 2022.
- [28] H. Son, J. W. Lim, S. Park, B. Park, J. Han, H. B. Kim, M. C. Lee, K. J. Jang, G. Kim, and J. H. Chung, "A Machine Learning Approach for the Classification of Falls and Activities of Daily Living in Agricultural Workers," *IEEE Access*, vol. 10, pp. 77418–77431, 2022.
- [29] G. Gao, Z. Li, Z. Huan, Y. Chen, J. Liang, B. Zhou, and C. Dong, "Human Behavior Recognition Model Based on Feature and Classifier Selection," *Sensors (Basel)*, vol. 21, no. 23, Dec. 2021.
- [30] F. M. Noori, M. Riegler, M. Z. Uddin, and J. Torresen, "Human Activity Recognition from Multiple Sensors Data Using Multi-fusion Representations and CNNs," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 16, no. 2, May 2020.
- [31] H. Hegazy, M. Abdelsalam, M. Hussien, S. Elmosalamy, Y. Hassan, A. Nabil, and A. Atia, "Multi-Sensor Fusion for Online Detection and Classification of Table Tennis Strokes," *International Journal of Intelligent Engineering and Systems*, vol. 14, no. 2, p. 2021.
- [32] Y. H. Nho, J. G. Lim, and D. S. Kwon, "Cluster-Analysis-Based User-Adaptive Fall Detection Using Fusion of Heart Rate Sensor and Accelerometer in a Wearable Device," *IEEE Access*, vol. 8, pp. 40389–40401, 2020.
- [33] S. S. Bangaru, C. Wang, S. A. Busam, and F. Aghazadeh, "ANN-based automated scaffold builder activity recognition through wearable EMG and IMU sensors," *Autom Constr*, vol. 126, p. 103653, Jun. 2021.
- [34] G. Tripathy and A. Sharaff, "AEGA: enhanced feature selection based on ANOVA and extended genetic algorithm for online customer review analysis," *Journal of Supercomputing*, vol. 79, no. 12, pp. 13180–13209, Aug. 2023.
- [35] C. Vong, T. Theptit, V. Watcharakonpipat, P. Chanchotisatien, and S. Laitrakun, "Comparison of Feature Selection and Classification for Human Activity and Fall Recognition using Smartphone Sensors," *2021 Joint 6th International Conference on Digital Arts, Media and Technology with 4th ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunication Engineering, ECTI DAMT and NCON 2021*, pp. 170–173, Mar. 2021.
- [36] K. Karagiannaki, A. Panousopoulou, and P. Tsakalides, "A benchmark study on feature selection for human activity recognition," *UbiComp 2016 Adjunct - Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 105–108, Sep. 2016.
-