Analisis Model Machine Learning Terhadap Pengenalan Aktifitas Manusia

Rizki Rino Pratama

Jurusan Teknik Informatika Universitas Nusa Putra Sukabumi, Indonesia E-Mail: rizki.pratama@nusaputra.ac.id

ABSTRAK

Aktivitas-aktivitas manusia diklasifikasikan dengan menggunakan sensor responsif dari gerakan manusia yang disebut pengguna. Karya ilmiah ini berfokus pada penggunaan model klasifikasi pendekatan pembelajaran mesin yang berbeda. Dalam penelitian ini, data yang digunakan diambil dari *open source* yang diklasifikasikan untuk mengenali aktivitas manusia di mana percobaannya telah dilakukan dengan sekelompok 30 sukarelawan dalam berbagai kelompok usia. Setiap orang melakukan enam aktivitas mengenakan smartphone di bagian pinggang. Dengan menggunakan *accelerometer* dan *gyroscope* yang tertanam, ditangkap akselerasi linear 3-aksial dan kecepatan sudut 3-aksial pada kecepatan konstan 50Hz. Dataset yang diperoleh telah dipartisi secara acak menjadi dua set, di mana 70% sukarelawan dipilih untuk menghasilkan data *training* dan 30% untuk data uji. Hasil pendekatan yang digunakan dibandingkan dalam hal efisiensi akurasi dan presisi. Model yang digunakan adalah regresi logistik, linear SVC, rbf SVM classifier, decision tree, dan random forest.

Kata kunci: Machine Learnig, Linear Regression, SVM, Decision Tree, Linear SVC.

ABSTRACT

Human activities classify the activity of a person using the sensor responsive affect from human movement which called users. Users are increasing the ability of prediction methods in machine learning. These facts make it more important and popular. This work focuses on using models of machine learning approach classification of different. In this research, the data used were taken from the open source classified to recognize human activity where the experiments have been carried out with a group of 30 volunteers within group an age. Each person performed six activities wearing a smartphone on the waist. Using its embedded accelerometer and gyroscope, we captured 3-axial linear acceleration and 3-axial angular velocity at a constant rate of 50Hz. The obtained dataset has been randomly partitioned into two sets, where 70% of the volunteers was selected for generating the training data and 30% the test data. The results of the approach used are compared in terms of efficiency prediction and precision. The models used is logistic regression, linear SVC, rbf SVM classifier, decision tree, and random forest.

Keywords: Machine Learning, Linear Regression, SVM, Decision Tree, Linear SVC.

Author Korespondensi (Rizki Rino Pratama) Email: rizki.pratama@nusaputra.ac.id

I. PENDAHULUAN

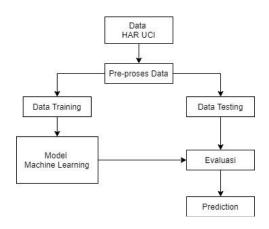
Pengakuan aktivitas manusia memainkan peran penting dalam interaksi manusia dan hubungan interpersonal. Untuk meningkatkan penelitian ini di bidang interaksi manusia telah diklasifikasikan kegiatan manusia dilakukan dengan menggunakan pembelajaran mesin. Pembelajaran mesin adalah ide cemerlang dalam dekade belakangan ini, karya ilmiah ini menerapkan beberapa model pembelajaran mesin untuk mendeteksi aktivitas manusia.

Penelitian ini menggunakan data yang sudah ada pada open source UCI yaitu HAR (Human Activities Recognition). Data tersebut diklasifikasikan untuk mengenali aktivitas manusia di mana percobaannya telah dilakukan dengan sekelompok 30 sukarelawan dalam berbagai kelompok usia. Setiap orang melakukan enam aktivitas mengenakan smartphone di bagian pinggang. Dengan menggunakan accelerometer dan gyroscope yang tertanam, ditangkap akselerasi linear 3-aksial dan kecepatan sudut 3-aksial pada kecepatan konstan 50Hz. Dataset yang diperoleh telah dipartisi secara acak menjadi dua set, di 70% sukarelawan dipilih mana menghasilkan data training dan 30% untuk data uji. Tujuannya adalah untuk mengklasifikasikan kegiatan menjadi salah satu dari enam kegiatan yang membuat visualisasi walking, walking upstairs, walking downstairs, sitting, standing, dan laving.

Karya ilmiah ini memiliki tujuan untuk menunjukkan bahwa algoritma pembelajaran mesin mampu memprediksi kegiatan-kegiatan manusia berdasarkan data yang sudah ada menjadi lebih baik selain itu dapat memprediksi secara akurat dengan model pembelajaran mesin. Karya ilmiah ini akan mulai dengan menganalisis data yang diambil dan kemudian menerapkan algoritma pembelajaran mesin.

II. METODOLOGI

Kerangka konsep didalam penelitian ini adalah sebagai berikut.



Gambar 1. Kerangka Konsep

Data yang dipakai dalam penelitian ini yaitu HAR UCI yang didaptakan dari repository machine learning UCI. Melalui tahapan pre-proses data, data yang ada akan dibagi menjadi dua dimana 70% data dipakai untuk data training dan sisa 30% akan digunakan untuk data uji.

Kemudian untuk model machine learning yang dipakai adalah Decision Tree, Random Forest, Support Vector Machine, Logistic Regression, Support-Vector Cluster. Setalah itu, evaluasi dengan melihat prediksi hasil pengenalan aktivitas manusianya.

2.1. Pembelajaran Mesin

Pembelajaran mesin (ML) adalah studi ilmiah tentang algoritma dan model statistik yang digunakan sistem komputer untuk melakukan tugas tertentu tanpa menggunakan instruksi eksplisit, dengan mengandalkan pola dan inferensi sebagai gantinya. Itu dikenal sebagai bagian dari kecerdasan buatan. Algoritma pembelajaran mesin membangun model matematika berdasarkan data sampel, yang dikenal sebagai "data training",

untuk membuat prediksi atau keputusan tanpa diprogram secara eksplisit untuk melakukan tugasnya [1]. Algoritma pembelajaran mesin digunakan dalam berbagai aplikasi, seperti penyaringan email dan visi komputer, di mana sulit atau tidak mungkin untuk mengembangkan algoritma konvensional untuk melakukan tugas secara efektif.

Pembelajaran mesin terkait erat dengan statistik komputasi, yang berfokus pada membuat prediksi menggunakan komputer. Studi tentang optimasi matematika memberikan metode, teori dan domain aplikasi ke bidang pembelajaran mesin. Penambangan data adalah bidang studi dalam pembelajaran mesin, dan berfokus pada analisis data eksplorasi melalui pembelajaran tanpa pengawasan [2].

2.2. Supervised Learning

Algoritma pembelajaran diawasi yang membangun model matematika dari serangkaian data yang berisi input dan output yang diinginkan [3]. Data tersebut dikenal sebagai data training, dan terdiri dari serangkaian contoh training. Setiap contoh training memiliki satu atau lebih input dan output yang diinginkan, juga dikenal sebagai sinyal pengawasan. Dalam model matematika, setiap contoh training diwakili oleh array atau vektor, kadang-kadang disebut vektor fitur, dan data training diwakili oleh sebuah matriks. Melalui optimalisasi berulang fungsi objektif, algoritma pembelajaran terawasi mempelajari fungsi yang dapat digunakan untuk memprediksi output yang terkait dengan input baru [4]. Fungsi yang optimal akan memungkinkan algoritma untuk menentukan output dengan benar untuk input yang bukan bagian dari data training. Algoritma yang meningkatkan akurasi output atau prediksi dari waktu ke waktu dikatakan telah belajar untuk melakukan tugas itu [5].

2.3. Unsupervised Learning

Unsupervised learning ienis adalah pembelajaran mesin yang mencari pola yang sebelumnya tidak terdeteksi dalam kumpulan data tanpa label yang sudah ada sebelumnya dan dengan minimal pengawasan manusia. Berbeda dengan pembelajaran terawasi yang biasanya menggunakan data berlabel pembelajaran tanpa pengawasan, juga dikenal sebagai swasusun memungkinkan pemodelan kepadatan probabilitas atas input [6]. Ini membentuk salah satu dari tiga kategori utama pembelajaran mesin, bersama dengan pembelajaran yang diawasi dan diperkuat. Pembelajaran semi-diawasi, varian terkait, memanfaatkan teknik yang diawasi dan tidak diawasi.

Dua metode utama yang digunakan dalam pembelajaran tanpa pengawasan adalah komponen utama dan analisis kluster. Analisis Cluster digunakan dalam pembelajaran tanpa pengawasan untuk mengelompokkan, atau mengelompokkan, kumpulan data dengan atribut bersama untuk mengekstrapolasi hubungan algoritmik Analisis cluster adalah cabang pembelajaran mesin vang mengelompokkan data yang belum diberi label, diklasifikasikan atau dikategorikan. Alihalih menanggapi umpan balik, analisis klaster mengidentifikasi kesamaan dalam data dan bereaksi berdasarkan ada tidaknya kesamaan di setiap potongan data baru. Pendekatan ini membantu mendeteksi titik data anomali yang tidak cocok dengan kelompok mana pun.

2.4. Semi-Supervised Learning

Semi-Supervised Learning adalah pendekatan pembelajaran mesin yang menggabungkan sejumlah kecil data berlabel dengan sejumlah besar data yang tidak berlabel selama training. Pembelajaran semi-diawasi berada di antara pembelajaran tanpa pengawasan (tanpa data training yang berlabel) dan pembelajaran yang diawasi (dengan hanya data training yang berlabel) [7].

2.5. Model

Melakukan pembelajaran mesin melibatkan pembuatan model, yang dilatih pada beberapa data pelatihan dan kemudian dapat memproses data tambahan untuk membuat prediksi. Berbagai jenis model telah digunakan dan diteliti untuk sistem pembelajaran mesin.

2.5.1. Logistic Regression

Dalam statistik, regresi logistik adalah pendekatan linier untuk memodelkan hubungan antara respons skalar (atau variabel dependen) dan satu atau lebih variabel penjelas (atau variabel independen). Kasus satu variabel penjelas disebut regresi linier sederhana. Untuk lebih dari satu variabel penjelas, prosesnya disebut regresi linier berganda [8]. Istilah ini berbeda dari regresi linier multivariat, di mana beberapa variabel dependen

berkorelasi diprediksi, daripada variabel skalar tunggal.

Dalam regresi linier, hubungan dimodelkan menggunakan fungsi prediktor linier yang parameter model tidak diketahui diperkirakan dari data. Model seperti ini disebut model linier [9]. Paling umum, rata-rata kondisional dari respons yang diberikan nilai-nilai dari variabel penjelas (atau prediktor) diasumsikan sebagai fungsi affine dari nilai-nilai tersebut; lebih jarang, median kondisional atau kuantil lainnya digunakan. Seperti semua bentuk analisis regresi, regresi linier berfokus pada distribusi probabilitas bersyarat dari respons yang diberikan nilai-nilai prediktor, daripada pada distribusi probabilitas gabungan dari semua variabel ini, yang merupakan domain dari analisis multivariat.

$$y = \beta x + \varepsilon$$

(1)

Dimana,

y adalah vektor dari nilai yang diamati

 $y_i(i = 1, 2, ..., n)$

β adalah vektor parameter dimensi (p+1) x matriks baris vektor $x_i(i = 1,2,...,n)$

ε adalah vektor nilai ε_i

2.5.2. Support Vector Machines

Dalam pembelajaran mesin, mesin dukunganvektor (SVM, juga jaringan dukungan-vektor [10]) adalah model pembelajaran yang diawasi dengan algoritma pembelajaran terkait yang menganalisis data yang digunakan untuk klasifikasi dan analisis regresi. Diberikan serangkaian contoh pelatihan, masing-masing ditandai sebagai milik satu atau yang lain dari dua kategori, algoritma pelatihan SVM membangun model yang memberikan contoh baru untuk satu kategori atau yang lain, menjadikannya sebagai pengelompokan linear biner non-probabilistik (walaupun metode seperti skala Platt ada untuk menggunakan SVM dalam pengaturan klasifikasi probabilistik). Model SVM adalah representasi dari contoh-contoh sebagai titik-titik dalam ruang, dipetakan sehingga contohcontoh kategori yang terpisah dibagi oleh celah yang jelas yang selebar mungkin. Contoh-contoh baru kemudian dipetakan ke dalam ruang yang sama dan diprediksi termasuk dalam kategori berdasarkan sisi kesenjangan di mana mereka iatuh.

Selain melakukan klasifikasi linier, SVM dapat secara efisien melakukan klasifikasi nonlinier menggunakan apa yang disebut trik kernel, secara implisit memetakan input mereka ke dalam ruang fitur dimensi tinggi.

Dalam program penelitian ini menggunakan SVC linear (support-vector clustering) dan rbf (radial basis function) classifier SVM.

$$\vec{\omega} \cdot \vec{x} - b = 0 \tag{2}$$

dimana $\vec{\omega}$ adalah normal vektor .

2.5.3. Decision Tree Learning

Pembelajaran pohon keputusan adalah metode yang umum digunakan dalam penambangan data [11]. Tujuannya adalah untuk membuat model yang memprediksi nilai variabel target berdasarkan beberapa variabel input. Pohon keputusan adalah representasi sederhana untuk mengklasifikasikan contoh. Untuk bagian ini, asumsikan bahwa semua fitur input memiliki domain diskrit terbatas, dan ada fitur target tunggal yang disebut "klasifikasi". Setiap elemen dari domain klasifikasi disebut kelas. Pohon keputusan atau pohon klasifikasi adalah pohon di mana setiap simpul internal (non-daun) diberi label dengan fitur input. Busur yang berasal dari simpul yang ditandai dengan fitur input diberi label dengan masing-masing nilai yang mungkin dari fitur target atau output atau busur mengarah ke simpul keputusan bawahan pada fitur input vang berbeda. Setiap daun pohon diberi label dengan kelas atau distribusi probabilitas di atas kelas, menandakan bahwa kumpulan data telah diklasifikasikan oleh pohon ke dalam kelas tertentu, atau ke dalam distribusi probabilitas tertentu.

$$H(T) = -\sum_{i=1}^{J} p_i \log_2 p_i$$

(3)

Dimana,

H(T) adalah entropi

 P_i informasi gain (i=1,2,...,n)

Dalam program penelitian ini menggunakan pohon keputusan dan random forest.

III. HASIL DAN PEMBAHASAN

Pada bab ini peniliti akan memaparkan hasil dari penilitian yang sudah dilaksanakan, berikut adalah penjelasannya.

3.1. Data

Data yang dipakai dalam penelitian ini yaitu HAR UCI yang didaptakan dari repository machine learning UCI. Melalui tahapan pre-proses data, data yang ada akan dibagi menjadi dua dimana 70% data dipakai untuk data training dan sisa 30% akan digunakan untuk data uji. Gambar 2 mendeskripsikan data training yang akan dipakai dalam model machine learning (bagian awal). Sedangkan Gambar 3 mendeskripsikan data training yang akan dipakai dalam model machine learning (bagian akhir).

In [27]:	train										
Out[27]:		tBodyAccmeanX	tBodyAccmeanY	tBodyAccmeanZ	tBodyAccstdX	tBodyAccstdY	tBodyAccstdZ	tBodyAccmadX	tBodyAccmadY	tBodyAccmadZ	tBoo
	0	0.288585	-0.020294	-0.132905	-0.995279	-0.983111	-0.913526	-0.995112	-0.983185	-0.923527	
	1	0.278419	-0.016411	-0.123520	-0.998245	-0.975300	-0.960322	-0.998807	-0.974914	-0.957686	
	2	0.279653	-0.019467	-0.113462	-0.995380	-0.967187	-0.978944	-0.996520	-0.963668	-0.977469	
	3	0.279174	-0.026201	-0.123283	-0.996091	-0.983403	-0.990675	-0.997099	-0.982750	-0.989302	
	4	0.276629	-0.016570	-0.115362	-0.998139	-0.980817	-0.990482	-0.998321	-0.979672	-0.990441	
	7347	0.299665	-0.057193	-0.181233	-0.195387	0.039905	0.077078	-0.282301	0.043616	0.060410	
	7348	0.273853	-0.007749	-0.147468	-0.235309	0.004816	0.059280	-0.322552	-0.029456	0.080585	
	7349	0.273387	-0.017011	-0.045022	-0.218218	-0.103822	0.274533	-0.304515	-0.098913	0.332584	
	7350	0.289654	-0.018843	-0.158281	-0.219139	-0.111412	0.268893	-0.310487	-0.068200	0.319473	
	7351	0.351503	-0.012423	-0.203867	-0.269270	-0.087212	0.177404	-0.377404	-0.038678	0.229430	

Gambar 2. Data Train (begin feature)

	train							
27]:		angletBodyGyroJerkMeangravityMean	angleXgravityMean	angleYgravityMean	angle Z gravityMean	subject	Activity	ActivityName
	-0.464761	-0.018446	-0.841247	0.179941	-0.058627	1	5	STANDING
	-0.732626	0.703511	-0.844788	0.180289	-0.054317	1	5	STANDING
	0.100699	0.808529	-0.848933	0.180637	-0.049118	1	5	STANDING
	0.640011	-0.485366	-0.848649	0.181935	-0.047663	1	5	STANDING
	0.693578	-0.615971	-0.847865	0.185151	-0.043892	1	5	STANDING
	0.206972	-0.425619	-0.791883	0.238604	0.049819	30	2	WALKING_UPSTAIRS
	-0.879033	0.400219	-0.771840	0.252676	0.050053	30	2	WALKING_UPSTAIRS
	0.864404	0.701169	-0.779133	0.249145	0.040811	30	2	WALKING_UPSTAIRS
	0.936674	-0.589479	-0.785181	0.246432	0.025339	30	2	WALKING_UPSTAIRS
	-0.056088	-0.616956	-0.783267	0.246809	0.036695	30	2	WALKING_UPSTAIR

Gambar 3. Data Train (end feature)

3.2. Data Pre-Processing

Pra-pemrosesan data adalah salah satu langkah terpenting dalam pembelajaran mesin, yang membantu dalam membangun model pembelajaran mesin dengan lebih akurat. Setiap data scientist harus menghabiskan 80% waktu untuk pra-pemrosesan data dan 20% waktu untuk benar-benar melakukan analisis. Pra-pemrosesan data adalah proses pembersihan data mentah. Data dikumpulkan dalam data raw dan dikonversi ke set data bersih. Dengan kata lain, setiap kali data dikumpulkan dari sumber yang berbeda dikumpulkan dalam format raw dan data ini tidak layak untuk analisis. Oleh karena itu, langkahlangkah tertentu dijalankan untuk mengubah data menjadi kumpulan data kecil yang bersih, bagian dari proses ini disebut data pra-pemrosesan.

3.3. Model

Sistem AI disajikan dengan data yang diberi label, itu berarti bahwa setiap data ditandai dengan label yang benar. Semi-supervised learning dikategorikan ke dalam tiga kategori yaitu klasifikasi, Regresi, dan pengelompokan. Masalah klasifikasi adalah ketika variabel target kategorikal (berlabel). Sedangkan masalah regresi adalah ketika variabel target kontinu, dan masalah pengelompokan adalah ketika variabel target dapat membuat beberapa cluster.

Di bawah ini ada beberapa klasifikasi, regresi dan algoritma klaster yang paling banyak digunakan.

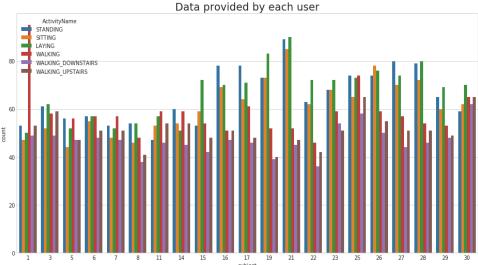
- 1. Decision Tree
- 2. Random Forest
- 3. Support Vector Machine
- 4. Logistic Regression
- 5. Support-Vector Cluster

Dari list yang sudah ada diatas, maka model yang dipakai dalam penilitian ini adalah kelima model yang sudah disebutkan diatas.

3.4. Training and testing model Machine Learning

Dari hasil eksplorasi data yang sudah dikerjakan menghasilkan sebagai berikut.





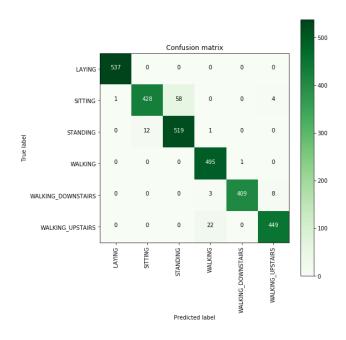
Gambar 4. Data Provided dari user

Gambar 4 mendeskripsikan data provided yang dihasilkan oleh semua pengguna yang ada dalam dataset HAR UCI.

Dengan menggunakan data provided diatas peneliti menghasilkan confusion matrix dan normalized confusion matrix yang akan mendukung prediksi pengenalan aktivitas manusia nantinya.

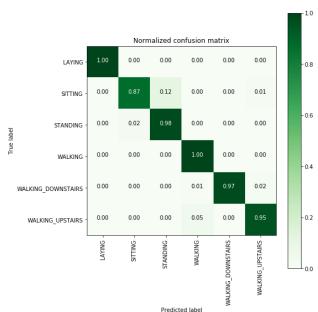
Hasil confusion matrix dapat dilihat pada gambar 5 dimana confusion matrix belum normal.

Kemudian pada gambar 6 adalah confusion matrix yang sudah normal.



Gambar 5. Confusion Matrix

Gambar diatas merupakan *confusion matrix* yang belum normal.



Gambar 6. Normalized Confusion Matrix

Gambar diatas merupakan confusion matrix yang sudah normal.

3.5. Evaluation

Tahap ini adalah tahap terakhir dalam bab ini dimana pada tahap ini adalah hasil proses model yang sudah dikomputasi dengan output adalah prediksi dan presisi dari berbagai model yang sudah dijelaskan sebelumnya di tahap sebelumnya. Berikut adalah hasil dari prediksi dan presisi yang dihasilkan dari berbagai model yang sudah dijelaskan di tahap sebelumnya.

3.5.1. Logistic Regression

Model pertama yang diujikan adalah model regresi logistik. Dimana akurasi yang dihasilkan adalah 0,96301 dan berikut hasil confusion matrix baru yang dihasilkan model regresi logistic dimana confusion matrix akan mempengaruhi presisi dan prediksi aktivitas manusia yang terbaca oleh modelnya.

		Acc	ura	 су	 	
	0.	9636	31323	33797	70818	3
				 -+ni		
				atri)		
[[:	537	6	9 6	9 6	9 6	0]
[2	428	57	0	0	4]
[0	11	520	1	0	0]
Ĩ	0	0	0	495	1	0]
Ĩ	0	0	0	3	409	8]
Ē	0	0	0	22	0	44911

Classifiction Rep	ort			
	precision	recall	f1-score	support
LAYING	1.00	1.00	1.00	537
SITTING	0.97	0.87	0.92	491
STANDING	0.90	0.98	0.94	532
WALKING	0.95	1.00	0.97	496
WALKING_DOWNSTAIRS	1.00	0.97	0.99	420
WALKING_UPSTAIRS	0.97	0.95	0.96	471
accuracy			0.96	2947
macro avg	0.97	0.96	0.96	2947

0.96

Gambar. 7. Hasil Logistic Regression

0.96

0.96

2947

Data yang didapat dalam gambar 7 dalam bentuk decimal. Dimana nilai akurasi yang di dapat adalah 0,96 dengan hasil flooring (pembulatan kebawah).

Model kedua yang diujikan adalah model Linear SVC. Dimana akurasi yang dihasilkan adalah 0,9653885 dan berikut hasil confusion matrix baru yang dihasilkan model Linear SVC dimana confusion matrix akan mempengaruhi presisi dan prediksi aktivitas manusia yang terbaca oleh modelnya.

		Accı	ırad	у		
	0.9	96538	8853	30709	91958	3
C	onfi	usior	n Ma	atri)	·	
[[537	0	6	9 6	9 6	9 9]
[2 4	121	63	0	0	5]
[0	10 5	21	1	0	0]
Ī	0	0	0	495	0	1]
Ī	0	0	0	2	413	5]
Ī	0	0	0	12	1	458]]

Classifiction Rep	ort			
	precision	recall	f1-score	support
LAYING	1.00	1.00	1.00	537
SITTING	0.98	0.86	0.91	491
STANDING	0.89	0.98	0.93	532
WALKING	0.97	1.00	0.98	496
WALKING_DOWNSTAIRS	1.00	0.98	0.99	420
WALKING_UPSTAIRS	0.98	0.97	0.97	471
accuracy			0.97	2947
macro avg	0.97	0.97	0.97	2947
weighted avg	0.97	0.97	0.97	2947

Gambar 8. Hasil Linear SVC

Data yang didapat dalam gambar 8 dalam bentuk decimal. Dimana nilai akurasi yang di adalah 0,97 dapat dengan hasil ceiling (pembulatan keatas).

3.5.2. Linear SVC

weighted avg

3.5.3. RBF SVM Classifier

Model ketiga yang diujikan adalah model *RBF SVM Classifier*. Dimana akurasi yang dihasilkan adalah 0,96267 dan berikut hasil confusion matrix baru yang dihasilkan model *RBF SVM Classifier* dimana confusion matrix akan mempengaruhi presisi dan prediksi aktivitas manusia yang terbaca oleh modelnya.

Classifiction Report									
	precision	recall	f1-score	support					
LAYING	1.00	1.00	1.00	537					
SITTING	0.97	0.90	0.93	491					
STANDING	0.92	0.98	0.95	532					
WALKING	0.96	0.99	0.97	496					
WALKING_DOWNSTAIRS	0.99	0.95	0.97	420					
WALKING_UPSTAIRS	0.95	0.96	0.95	471					
accuracy			0.96	2947					
macro avg	0.96	0.96	0.96	2947					
weighted avg	0.96	0.96	0.96	2947					

Gambar 9. Hasil rbf SVM Classifier

Data yang didapat dalam gambar 9 dalam bentuk decimal. Dimana nilai akurasi yang di dapat adalah 0,96 dengan hasil flooring (pembulatan kebawah).

3.5.4. Decision Tree

Model keempat yang diujikan adalah model *Decision Tree*. Dimana akurasi yang dihasilkan adalah 0,8629 dan berikut hasil confusion matrix

baru yang dihasilkan model *Decision Tree* dimana confusion matrix akan mempengaruhi presisi dan prediksi aktivitas manusia yang terbaca oleh modelnya.

	1		^ ^ ^				I
	I		ACC	cura	2 y		
				1111		70011	,
		6.8	3625	11141	0000	79912	2
	1.6	c.		M			
	C	onti	ısıc	on Ma	atri	X	
			,	. ,			
					9 (
	L	0 3	885	106	0	0	0]
	[0	93	439	0	0	0]
	[0	0	0	471	17	8]
	[0	0	0	17	342	61]
	Ī	0	0	0	73	29	369]]
	-						
Classif	fiction	Report	:				
		pr	ecisi	on r	ecall	f1-score	support

	precision	recall	f1-score	support
LAYING	1.00	1.00	1.00	537
SITTING	0.81	0.78	0.79	491
STANDING	0.81	0.83	0.82	532
WALKING	0.84	0.95	0.89	496
WALKING_DOWNSTAIRS	0.88	0.81	0.85	420
WALKING_UPSTAIRS	0.84	0.78	0.81	471
accuracy			0.86	2947
macro avg	0.86	0.86	0.86	2947
weighted avg	0.86	0.86	0.86	2947

Gambar 10. Hasil Decision Tree

Data yang didapat dalam gambar 10 dalam bentuk decimal. Dimana nilai akurasi yang di dapat adalah 0,86 dengan hasil flooring (pembulatan kebawah).

3.5.5. Random Forest

Model terakhir yang diujikan adalah model *Random Forest*. Dimana akurasi yang dihasilkan adalah 0,8629 dan berikut hasil confusion matrix baru yang dihasilkan model *Random Forest* dimana confusion matrix akan mempengaruhi

presisi dan prediksi aktivitas manusia yang terbaca oleh modelnya.

Accuracy	

0.9110960298608755

Confusion	Matrix

[[537	7 6	9 6	9 6	9 6	9 0]
[0	417	74	0	0	0]
[0	44	488	0	0	0]
[0	0	0	484	10	2]
[0	0	0	35	339	46]
[0	0	0	45	6	420]]

Classifiction	Report

	precision	recall	f1-score	support
LAYING	1.00	1.00	1.00	537
SITTING	0.90	0.85	0.88	491
STANDING	0.87	0.92	0.89	532
WALKING	0.86	0.98	0.91	496
WALKING_DOWNSTAIRS	0.95	0.81	0.87	420
WALKING_UPSTAIRS	0.90	0.89	0.89	471
accuracy			0.91	2947
macro avg	0.91	0.91	0.91	2947
weighted avg	0.91	0.91	0.91	2947

Gambar 11. Result Random Forest

Data yang didapat dalam gambar 11 dalam bentuk decimal. Dimana nilai akurasi yang di dapat adalah 0,91 dengan hasil flooring (pembulatan kebawah).

	Accuracy	Error
Logistic Regression	: 96.3%	3.699%
Linear SVC	: 96.54%	3.461%
rbf SVM classifier	: 96.27%	3.733%
DecisionTree	: 86.29%	13.71%
Random Forest	: 91.11%	8.89%

Gambar 12. Hasil semua model *Machine Learning*

Gambar 12 mendeskripsikan hasil dari semua model yang diujikan dalam bentuk decimal. Dimana nilai akurasi dan error yang dihasilkan ada dalam tabel 1.

Tabel 1. Machine Learning models.

Models	Accuracy	Error
Logistic Regression	96,3 %	3,699 %
Linear SVC	96,54 %	3,461 %
rbf SVM classifier	96,27 %	3,733 %
Decision Tree	86, 29 %	13,71 %
Random Forest	91,11 %	8,89 %

Sesuai hasil yang ada dalam tabel 1 mendeskripsikan bahwa model-model yang diujikan dalam pengenalan aktivitas manusia menghasilkan nilai yang baik dimana hasilnya diatas 80%.

Model *Linear SVC* menghasilkan 96,54 % yang artinya bahwa diantara lima model yang diujikan model ini adalah model dengan hasil akurasi tertinggi, sedangkan model *decision tree* menghasilkan akurasi 86,29 % yaitu akurasi terendah diantara model-model lainnya.

Aktivitas manusia yang dujikan terdiri dari enam kategori yaitu, walking, walking upstairs, walking downstairs, sitting, standing, dan laying. Dimana hasil prediksi tertinggi yang di hasilkan adalah 1 untuk pengenalan aktivitas laying dengan menggunakan semua model machine learning.

IV. SIMPULAN DAN SARAN

Berdasarkan hasil uji yang telah dilakukan dengan lima model yang berbeda, maka menghasilkan nilai akurasi yang mempunyai tingkat error terendah adalah dengan menggunakan *Linear SVC* sebesar 96,54 % dan error 3,461 %, dengan nilai prediksi untuk walking, walking upstairs, walking downstairs, sitting, standing, dan laying berturut turut sebesar

1, 0.91, 0.93, 0.98, 0.99, 0.97 dalam satuan desimal.

Proses penentuan parameter merupakan bagian tersulit dalam tahapan perhitungan model machine learning karena terdapat beberapa parameter yang dikombinasikan. Pada model decision tree menghasilkan akurasi 86,29 % yaitu akurasi terendah diantara model-model lainnya. Perlu adanya kajian ulang terkait parameter-parameter dan model lain yang tepat sehingga hasil prediksi dan akurasi pengenalan aktivitas manusia semakin tinggi. Selain dikaji lagi dengan penggunan optimasi meta-heuristic atau optimasi numerik lainnya dalam penekanan persentase error dan hasil presisinya.

V. UCAPAN TERIMAKASIH

Ucapan terimakasih dihaturkan kepada keluarga yang telah mendukung dan Universitas Nusa Putra Sukabumi yang telah memberi kesempatan waktu bagi penulis untuk menyelesaikan penelitian ini.

REFERENSI

- [1] C. M. Bishop, *Pattern Recognition and Machine Learning*. 2006.
- [2] J. H. Friedman, "Data Mining and Statistics: What's the connection?," *Statistics (Ber).*, 1997.

- [3] S. Russell dan P. Norvig, Artificial Intelligence A Modern Approach Third Edition. 2010.
- [4] M. Mohri, A. Rostamizadeh, dan A. Talwalkar, Foundations of Machine Learning (Adaptive Computation and Machine Learning series). 2012.
- [5] T. Mitchell, *Machine Learning*. McGraw Hill, 1997.
- [6] V. Roman, "Unsupervised Machine Learning: Clustering Analysis -- Towards Data Science," *Towar. Data Sci.*, 2019.
- [7] D. A. Freedman, *Statistical models: Theory and practice*. 2009.
- [8] H. J. Scudder, "Probability of Error of Some Adaptive Pattern-Recognition Machines," *IEEE Transactions on Information Theory*. 1965, doi: 10.1109/TIT.1965.1053799.
- [9] H. L. Seal, "Studies in the History of Probability and Statistics. XV: The Historical Development of the Gauss Linear Model," *Biometrika*, 1967, doi: 10.2307/2333849.
- [10] C. Cortes dan V. Vapnik, "Support-Vector Networks," *Mach. Learn.*, 1995, doi: 10.1023/A:1022627411411.
- [11] L. Rokach dan O. Maimon, *Data mining with decision trees: theroy and applications*. 2008.