ML – PROJECT Report

ML-25-15 Activity Recognition Using the Smartphone

Nouman NTELI IMPRAIM -- 030720126 Muhammed Emin ASLANTEPE - 030721034

Overview

Overview	<i>2</i>
Introduction	3
Dataset and Preprocessing	3
Models and Methods	4
Evaluation and Results	4
Figures	5
Reference	<i>7</i>

Introduction

With the increasing prevalence of smartphones equipped with inertial sensors, it is now possible to collect data about human physical activities in real-world settings. Recognizing such activities through sensor data has gained importance in various domains such as healthcare, fitness tracking, elderly monitoring, and personal assistance systems.

This project aims to classify six daily human activities—WALKING, WALKING_UPSTAIRS, WALKING_DOWNSTAIRS, SITTING, STANDING, and LAYING—based on accelerometer and gyroscope data collected from a smartphone worn at the waist. By leveraging machine learning algorithms, we developed a system that can accurately predict the activity being performed using pre-processed feature vectors derived from time-series data

Dataset and Preprocessing

The dataset used in this study is the "Simplified Human Activity Recognition Using Smartphones" dataset published on Kaggle. It is based on sensor data collected from 30 individuals aged 19 to 48 years while performing six daily activities. Each subject wore a Samsung Galaxy S II smartphone on the waist, which recorded 3-axial linear acceleration and angular velocity at a constant sampling rate of 50Hz.

The raw signals were pre-processed using Butterworth filters and segmented into fixed-width sliding windows of 2.56 seconds with 50% overlap, resulting in 128 readings per window. From each segment, a total of 561 features were extracted based on time and frequency domain signal characteristics.

The dataset was split into a training set (80%) and a validation set (20%). Standard scaling ('StandardScaler') was applied to normalize feature values.

Models and Methods

Three machine learning algorithms were implemented and compared using a standardized pipeline that included feature scaling:

Logistic Regression: A linear model suitable for multiclass classification and interpretable results.

Random Forest: An ensemble learning method using multiple decision trees with majority voting. It handles overfitting better and can rank feature importance.

Support Vector Machine (SVM): A robust classifier effective in high-dimensional spaces and widely used in activity recognition tasks.

Each model was trained using stratified cross-validation and evaluated on the same validation set.

Evaluation and Results

Each model was evaluated using classification metrics including accuracy, precision, recall, and F1-score. Below is the comparison table of the three models:

	Model	F1-macro	Accuracy
0	logreg	0.9638	0.9626
1	rf	0.9580	0.9598
2	svm	0.9580	0.9571

The confusion matrix for Logistic Regression is shown below:

precision	recall	f1-score	support
0.99	1.00	1.00	136
0.95	0.87	0.91	125
0.90	0.96	0.92	134
0.98	1.00	0.99	121
0.99	0.98	0.98	98
0.99	0.97	0.98	108
		0.96	722
0.97	0.96	0.96	722
0.96	0.96	0.96	722
	0.99 0.95 0.90 0.98 0.99 0.99	0.99 1.00 0.95 0.87 0.90 0.96 0.98 1.00 0.99 0.98 0.99 0.97	0.99 1.00 1.00 0.95 0.87 0.91 0.90 0.96 0.92 0.98 1.00 0.99 0.99 0.98 0.98 0.99 0.97 0.98

The model performed best in distinguishing the LAYING, WALKING, and WALKING_UPSTAIRS classes. However, misclassifications were observed between SITTING and STANDING, which are naturally harder to differentiate due to similar postures.

The Logistic Regression model was selected as the best-performing model and saved as a `.joblib` file for future deployment.

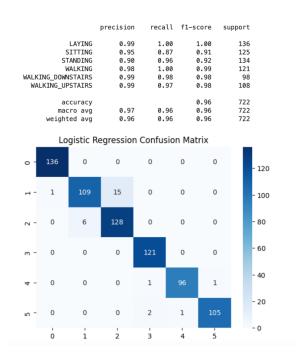
Conclusion

In this project, we implemented and compared three machine learning algorithms for human activity recognition using smartphone sensor data. All models achieved high accuracy (~96%), with Logistic Regression slightly outperforming the others in macro F1-score.

The results show that even simple models can be effective on this structured, feature-rich dataset. Future improvements may involve using raw signal input with deep learning architectures such as CNN or LSTM to capture temporal patterns more effectively.

Real-world applications include activity-based health tracking and mobile fitness apps.

Figures



pd.read_csv("../submission.csv").head()

SITTING							
LAYING 0.99 1.00 1.00 1 SITTING 0.95 0.87 0.91 1 STANDING 0.90 0.96 0.92 1 WALKING 0.98 1.00 0.99 1 WALKING_DOWNSTAIRS 0.99 0.98 0.98 WALKING_UPSTAIRS 0.99 0.97 0.98 1 accuracy	Mod	del:	logreg		11		
SITTING				precision	recall	T1-score	suppor
STANDING			LAYING	0.99	1.00	1.00	13
STANDING			SITTING	0.95	0.87	0.91	12
WALKING							13
WALKING_DOWNSTAIRS							12
### WALKING_UPSTAIRS 0.99 0.97 0.98 1 accuracy	WΔI	KTNG					
Macro avg 0.97 0.96 0.96 7							16
Macro avg 0.97 0.96 0.96 7			accuracy			0.96	72
Weighted avg 0.96 0.96 0.96 7				0.97	0.96		72
LAYING 0.99 1.00 1.00 1		W					72
LAYING 0.99 1.00 1.00 1	===						
LAYING 0.99 1.00 1.00 1 SITTING 0.95 0.94 0.95 1 STANDING 0.96 0.96 0.96 1 WALKING 0.97 0.98 0.98 1 WALKING_DOWNSTAIRS 0.94 0.94 0.93 1 WALKING_UPSTAIRS 0.94 0.93 0.93 1 accuracy 0.96 0.96 0.96 7 Weighted avg 0.96 0.96 0.96 7 Weighted avg 0.96 0.96 0.96 7 SITTING 0.99 0.99 0.99 1 SITTING 0.99 0.99 0.99 1 STANDING 0.90 0.93 0.92 1 STANDING 0.90 0.93 0.92 1 WALKING_DOWNSTAIRS 0.96 0.98 0.97 WALKING_UPSTAIRS 0.98 0.97 0.98 1 accuracy 0.96 0.96 0.96 0.96 7 WEIGHT 0.92 0.88 0.97 0.98 1 WALKING_UPSTAIRS 0.98 0.97 0.98 1 accuracy 0.96 0.98 0.97 WALKING_UPSTAIRS 0.98 0.97 0.98 1 ACCURACY 0.96 0.96 0.96 7 WE in iyi model kaydedildi: models/best_model.joblib Model Karşılaştırma: Model FI-macro Accuracy 0.96 0.96 7 Weighted avg 0.963 0.9626 1 rf 0.9580 0.9581 0.9571 V submission.csv başarıyla oluşturuldu! Id PredictedActivity 0 0 STANDING 1 1 STANDING	Mod	del:	rf		11	41	
SITTING				precision	recall	TI-score	suppo
STANDING 0.96 0.96 0.96 1			LAYING		1.00	1.00	13
STANDING 0.96 0.96 0.96 1			SITTING	0.95	0.94	0.95	12
WALKING_DOWNSTAIRS 0.94 0.94 0.93 1 accuracy 0.96 0.96 0.96 7 weighted avg 0.96 0.96 0.96 7 Bornell 1-score suppool 1-scor			STANDING		0.96	0.96	13
WALKING_DOWNSTAIRS 0.94 0.94 0.93 1 accuracy 0.96 0.96 0.96 7 weighted avg 0.96 0.96 0.96 7 Bornell 1-score suppool 1-scor			WALKING	0.97	0.98	0.98	12
### WALKING_UPSTAIRS 0.94 0.93 0.93 1 accuracy	WΔI	KTNG					-
accuracy							
macro avg 0.96 0.96 0.96 7	•	WENT	NO_OFSTAIRS	0.54	0.95	0.93	1,
Weighted avg 0.96 0.96 0.96 7							72
### Decision Precision Pr							72
Precision recall f1-score support		W	eighted avg	0.96	0.96	0.96	72
Precision recall f1-score support	===	===== del:	======================================				
SITTING				precision	recall	f1-score	suppor
SITTING			LAYING	0.99	0.99	0.99	13
WALKING			SITTING	0.92	0.88	0.90	1
WALKING			STANDING	0.90	0.93	0.92	1
WALKING_DOWNSTAIRS							
WALKING_UPSTAIRS 0.98 0.97 0.98 1 accuracy 0.96 7 macro avg 0.96 0.96 0.96 7 Weighted avg 0.96 0.96 0.96 7 ✓ En iyi model kaydedildi: models/best_model.joblib ii Model Karşılaştırma: Model FI-macro Accuracy 0 logreg 0.9638 0.9626 1 rf 0.9580 0.9598 2 svm 0.9580 0.9571 ✓ submission.csv başarıyla oluşturuldu! Id PredictedActivity 0 0 STANDING 1 1 STANDING 2 2 STANDING	WΔI	KTNG					-
macro avg 0.96 0.96 0.96 7 weighted avg 0.96 0.96 0.96 7 In iyi model kaydedildi: models/best_model.joblib Model Karşılaştırma: Model FI-macro Accuracy 0 logreg 0.9638 0.9626 1 rf 0.9580 0.9598 2 svm 0.9580 0.9571 3 submission.csv başarıyla oluşturuldu! Id PredictedActivity 0 0 STANDING 1 1 STANDING 2 2 STANDING							10
macro avg 0.96 0.96 0.96 7 weighted avg 0.96 0.96 0.96 7 In iyi model kaydedildi: models/best_model.joblib Model Karşılaştırma: Model FI-macro Accuracy 0 logreg 0.9638 0.9626 1 rf 0.9580 0.9598 2 svm 0.9580 0.9571 3 submission.csv başarıyla oluşturuldu! Id PredictedActivity 0 0 STANDING 1 1 STANDING 2 2 STANDING			accuracy			0.96	72
weighted avg 0.96 0.96 0.96 7 The first model kaydedildi: models/best_model.joblib Model Karsılaştırma: Model F1-macro Accuracy 0 logreg 0.9638 0.9626 1 rf 0.9580 0.9598 2 sym 0.9580 0.9571 Submission.csv başarıyla oluşturuldu! Id PredictedActivity 0 0 STANDING 1 1 STANDING 2 2 STANDING				0.96	0.96		72
Model Karşılaştırma: Model F1-macro Accuracy logreg		W					7
Model Karşılaştırma: Model F1-macro Accuracy logreg	$\overline{\checkmark}$	En i	yi model kay	dedildi: mod	dels/best_	model.jobl	ib
Model F1-macro Accuracy 0 logreg 0.9638 0.9626 1 rf 0.9558 0.9598 2 svm 0.9580 0.9571 ✓ submission.csv başarıyla oluşturuldu! Id PredictedActivity 0 0 STANDING 1 1 STANDING 2 2 STANDING		Mode	l Karcılactı	rma !			
0 logreg 0.9638 0.9626 1 rf 0.9580 0.9598 2 svm 0.9580 0.9571 ✓ submission.csv başarıyla oluşturuldu! Id PredictedActivity 0 0 STANDING 1 1 STANDING 2 2 STANDING							
1 rf 0.9580 0.9598 2 svm 0.9580 0.9571 vsubmission.csv başarıyla oluşturuldu! Id PredictedActivity 0 0 STANDING 1 1 STANDING 2 2 STANDING	а						
<pre>2 svm 0.9580 0.9571 vubmission.csv başarıyla oluşturuldu! Id PredictedActivity 0</pre>		cogi					
submission.csv başarıyla oluşturuldu! Id PredictedActivity 0 0 STANDING 1 1 STANDING 2 2 STANDING		_					
Id PredictedActivity 0 0 STANDING 1 1 STANDING 2 2 STANDING					iction ld.		
0 0 STANDING 1 1 STANDING 2 2 STANDING	V				ışturuldu:		
1 1 STANDING 2 2 STANDING	_						
2 2 STANDING	0						
	1	1	STANDII	NG			
3 3 STANDING	2	2	STANDII	NG			
	3	3	STANDII	NG			

4 4

STANDING

Reference

- [1] UCI Machine Learning Repository. Human Activity Recognition Using Smartphones Dataset.
- [2] Kaggle Dataset: Simplified HAR (Link)
- [3] Scikit-learn Documentation (Link)
- [4] Logistic Regression, Random Forest, and SVM implementation: sklearn
- [5] Butterworth Filtering for signal processing