Human Activity Recognition using Smartphone Sensor Data

Nelson Almeida G00412122

March 2025

1 Introduction

In this project, the aim is to classify different physical activities performed by human beings based on data obtained by smartphone accelerometers and gyroscope sensors that detect motion and device orientation. This project dataset includes readings from 30 volunteers performing six activities: Walking, Walking Upstairs, Walking Downstairs, Sitting, Standing, and Laying. The objective is to build and evaluate machine learning models for accurate activity classification.

2 Methods

2.1 Data Preprocessing

The dataset consists of 561 features derived from smartphone sensors (accelerometer and gyroscope). Before training the models, the following preprocessing steps were performed:

- **Handling Missing Values:** The dataset was evaluated for missing values. After ploting the heatmap we confirmed there was no missing data.

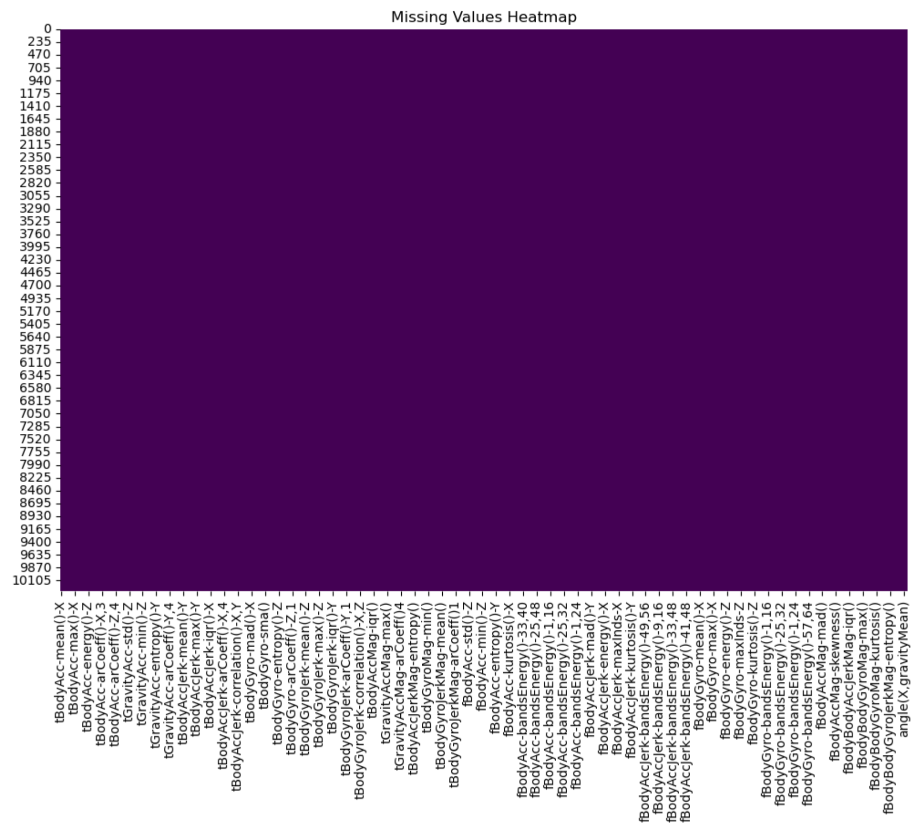

Figure 1: Caption

- **Class Imbalance:** The class distribution was analyzed using StandarScaler to normalize all feature values.
- **Feature Scaling:** StandarScaler was applied to normalize all feature values.
- **Encoding Labels:** Activity labels were mapped to numeric values (1-6) for model compatibility.
- **Feature Selection** Related features were removed using a correlation matrix analysis.

2.2 Model Training

Four machine learning models were trained and evaluated:

- Support Vector Machine (SVM) with linear and RBF kernels

- k-Nearest Neighbors (kNN)
- Random Forest
- Logistic Regression

The dataset was split into training and testing sets (70-30). Hyperameter tuning was performed using cross-validation to optimize model performance.

2.3 Model Evaluation

Models were compared using accuracy and classification reports. The best-performing model was selected based on:

- Test set accuracy
- Precision, really, and F1-score
- Confusion matrices to evaluate misclassifications

3 Results

3.1 Classifier Performance

Model	Acuraccy
SVM (Linear)	0.95
SVM (RBF)	**0.96**
kNN	0.92
Random Forest	0.94
Logistic Regression	0.93

Table 1: Accuracy scores of different classifiers

SVM with the RBF kernel performed the best with a 96% accuracy. The confusion matrix further revealed strong classification performance across all activity categories.

4 Discussion

SVM (RBF) outperformed other models due to its ability to capture complex patterns in sensor data. The use of standardized features contributed to improving the classifier stability. kNN had lower accuracy due to its sensitivity to large amounts of data, while Random Forest performed well but required longer training times.

5 Conclusion

In the end, the code successfully shows the application of various machine learning models for human activity recognition using smartphone sensor data. The results show that all models perform well, with SVM (RBF) achieving the highest accuracy of 96%. The confusion matrices and classification reports provide detailed information into the models' performance across different activities.

6 References

- Scikit-learn Documentation: `https://scikit-learn.org/stable/`
- Seaborn Documentation: `https://seaborn.pydata.org/`
- Matplotlib Documentation: `https://matplotlib.org/`
- Kaggle Documentation: `https://www.kaggle.com/code/kerneler/starter-traffic-driving-style`
- Sckit-learn Documentation:`https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression`
- Sckit-learn Documentation:`https://scikit-learn.org/stable/modules/preprocessing.html`
- GitHub Copilot: `https://github.com/features/copilot`