Walk vs Run Classification Project Report

# 1. Introduction

This project aims to classify whether a person is walking or running based on sensor data, including acceleration and gyroscope readings. The goal is to build a predictive model that can accurately identify the type of activity based on motion data.

# 2. Dataset Overview

The dataset contains 88,588 rows and 11 columns. The target variable is 'activity', where 0 indicates walking and 1 indicates running. Features include acceleration and gyroscope readings on three axes, along with a wrist identifier.

# 3. Data Cleaning and Preprocessing

Non-numeric columns such as 'date', 'time', and 'username' were dropped. There were no missing values or duplicate entries. The dataset was clean and suitable for analysis. The features were standardized using StandardScaler for model training.

# 4. Exploratory Data Analysis (EDA)

Correlation analysis revealed that 'acceleration\_y' had the strongest relationship with the activity label. Distributions and boxplots showed distinct patterns between walking and running, especially in the acceleration axes.

# 5. Feature Importance

Random Forest was used to evaluate feature importances. The top features were:  
- acceleration\_y (~46.7%)  
- acceleration\_z (~22.5%)  
- acceleration\_x (~17.1%)  
Gyroscope readings contributed less, and the wrist identifier had minimal impact.

# 6. Model Building and Evaluation

Three classification models were trained:  
- Logistic Regression  
- Random Forest  
- Support Vector Machine (SVM)  
  
Random Forest achieved the best performance with approximately 99.2% accuracy, while SVM followed closely. Logistic Regression performed relatively lower at 86.4% accuracy.

# 7. Real-Time Prediction

To classify a new observation:  
1. Gather sensor values for acceleration and gyroscope.  
2. Standardize the values using the fitted scaler.  
3. Use the trained Random Forest model to predict activity.  
4. Interpret the output: 0 = Walking, 1 = Running.

# 8. Conclusion

This project effectively demonstrates that motion sensor data can accurately differentiate walking from running. With a high-performing model, this system can be deployed in fitness trackers or mobile apps for real-time activity classification.