TRACKING BARBELL EXERCISES

by

Darshan Parsoliya

STB03007

Script. Sculpt. Socialize

Under the Guidance

of

Ms. Urooj Khan

Submitted to

Scifor Technologies

TABLE OF CONTENTS

ABSTRACT

This project focuses on advancing human activity recognition through sensor data analysis, specifically accelerometer and gyroscope readings. The primary objective is to develop a reliable system for accurately classifying various activities such as walking, running, and sitting. The project's methodology involves meticulous data preprocessing and cleaning, incorporating Chauvenet's Criterion for outlier detection, and replacing outliers with NaN values. Feature engineering plays a crucial role, encompassing diverse feature sets like basic, square, PCA-derived, temporal, frequency-based, and clustering features.
Techniques such as LowPass filtering, Principal Component Analysis (PCA), and Fourier Transformation are applied for feature extraction.

The project systematically evaluates different classification algorithms, including Neural Network, Random Forest, K-Nearest Neighbors, Decision Tree, and Naive Bayes. Grid searches for hyperparameter tuning and assessments on selected feature sets provide insights into model performance. Notably, Random Forest emerges as the top-performing model, highlighting the significance of feature selection and ensemble methods. The exploration extends to participant-specific evaluations, emphasizing the model's adaptability to individual variations.

In conclusion, this project establishes a robust framework for human activity recognition, covering critical aspects such as data preprocessing, feature engineering, model selection, and evaluation. The findings contribute valuable insights into the factors influencing classification accuracy, paving the way for practical applications in health monitoring and activity recognition systems.

PROBLEM STATEMENT

Human Activity Recognition (HAR) is a pivotal area in sensor data analysis, crucial for applications like healthcare monitoring, assistive technologies, and human-computer interaction. However, this field encounters multifaceted challenges. The diversity in activity patterns, inherent sensor noise, difficulties in real-world generalization, and capturing temporal dynamics present substantial hurdles. The project aims to overcome these challenges by enhancing the robustness of recognition models, addressing sensor noise issues, improving real-world generalization capabilities, and implementing mechanisms to effectively capture temporal dynamics

OBJECTIVE

- Enhance Sensor Fusion Techniques: Develop advanced sensor fusion techniques to effectively integrate information from Gyroscopes and Accelerometers, aiming to improve the accuracy and reliability of human activity recognition models.

- Optimize Recognition Algorithms: Implement and optimize machine learning algorithms tailored for Gyro and Accelo data, addressing specific challenges associated with these sensors to achieve superior activity recognition performance.

- Handle Sensor Noise and Variability: Devise strategies to mitigate noise and handle variability in sensor data, ensuring the robustness of the models across different users, devices, and environmental conditions.

- Real-time Recognition: Engineer the models for real-time human activity recognition, enabling timely responses and applications in dynamic, time-sensitive scenarios.

- Cross-Device Compatibility: Ensure compatibility and effectiveness across a variety of devices, facilitating seamless integration into diverse technology ecosystems

TOOLS & TECHNOLOGY

.

Software Requirement:

● Visual Studio Code

Technology:

1. Python
2. Machine Learning
3. Python libraries - Matplotlib,sklearn,pandas,scipy
4. Statistics
5. Data Visualization
6. Handling Data

ABOUT DATASET

Data Source:

1. Utilizing MetaMotion sensors for data collection.
2. Sensors include accelerometer and gyroscope modules.

Sensor Outputs:

1. Accelerometer: Captures acceleration forces experienced by the sensor.
2. Gyroscope: Measures the rate of rotation around different axes.

Dataset Information :

1. Data collected from 5 individuals.
2. Names of participants involved in the study.
3. Labels assigned to each exercise (e.g., bench press, over head press, deadlift).
4. Categorization based on exercise intensity (e.g., heavy, medium).
5. Data recorded over 3 axes (x, y, z) for both accelerometer and gyroscope.

DATA PREPROCESSING

Import & Organization:

1. Loaded MetaMotion sensor data (accelerometer & gyroscope) using Pandas.
2. Organized data over 3 axes (x, y, z) for both sensors.

Participants & Labels:

- Gathered data from 5 individuals.
- Extracted participant names, exercise labels, and conditions from file names.

Cleaning Steps:

- Removed unnecessary columns: 'epoch (ms)', 'time (01:00)', 'elapsed (s)'.
- Converted timestamp to datetime format for indexing.

Resampling Technique:

- Applied resampling at a consistent 200ms interval.
- Utilized mean for numerical data and last value for categorical data.
- Checked and addressed missing values in the resampled dataset.

Concatenation & Overview:

- Concatenated accelerometer and gyroscope data.
- Explored resulting dataset with combined information.

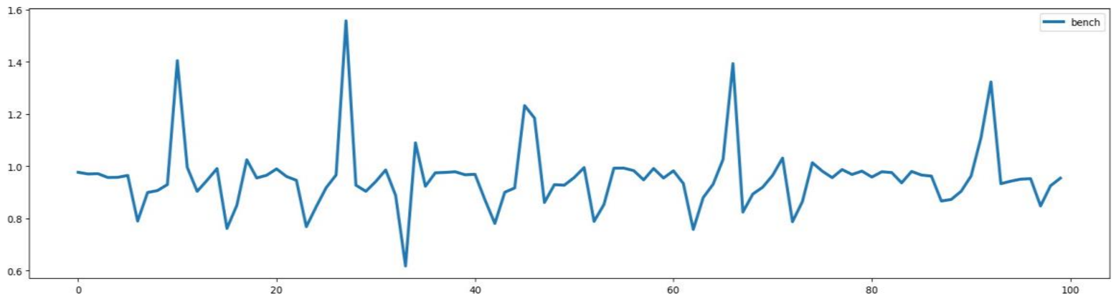
EXPLORATORY DATA ANALYSIS

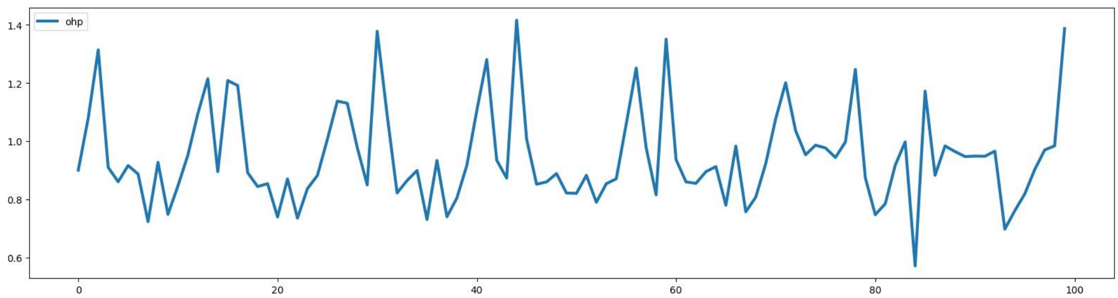
Exercise-specific Visuals:

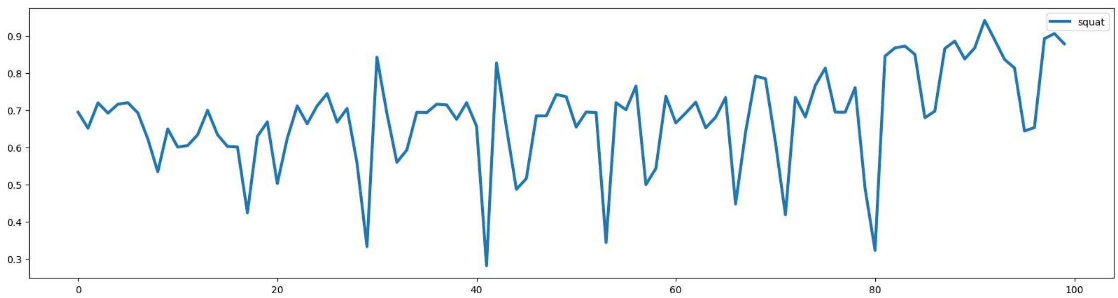
- Delved into individual exercises, providing dedicated plots for each exercise label.
- Focused on the 'acc_y' axis to capture exercise-specific patterns.

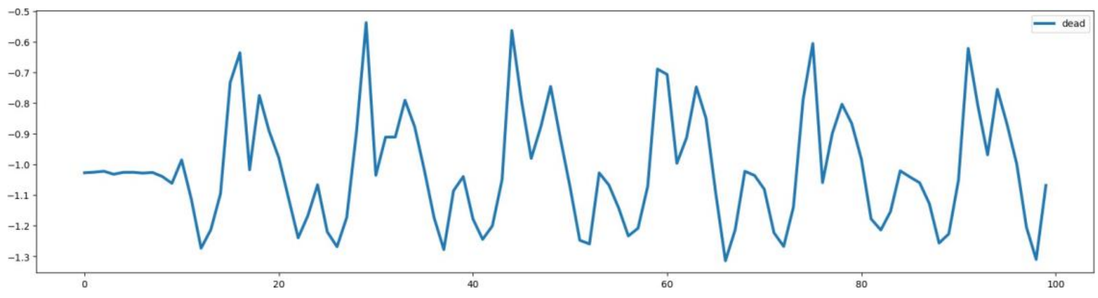
Subset Visualization (Limited Samples):

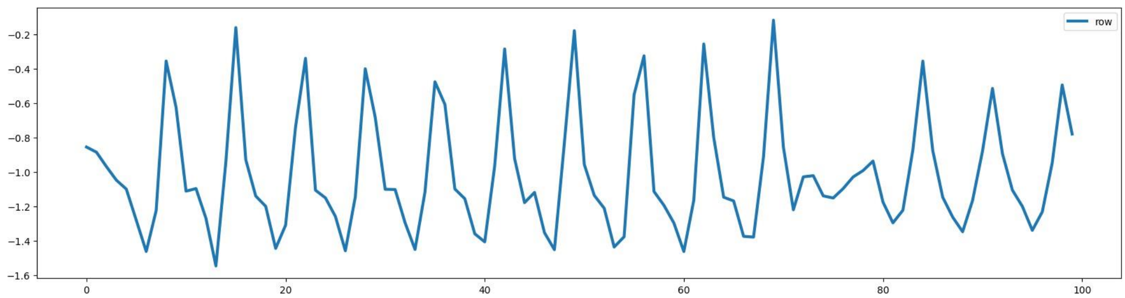
- Displayed a snapshot of the first 100 samples for a more detailed view.
- Concentrated on the 'acc_y' axis to observe variations.

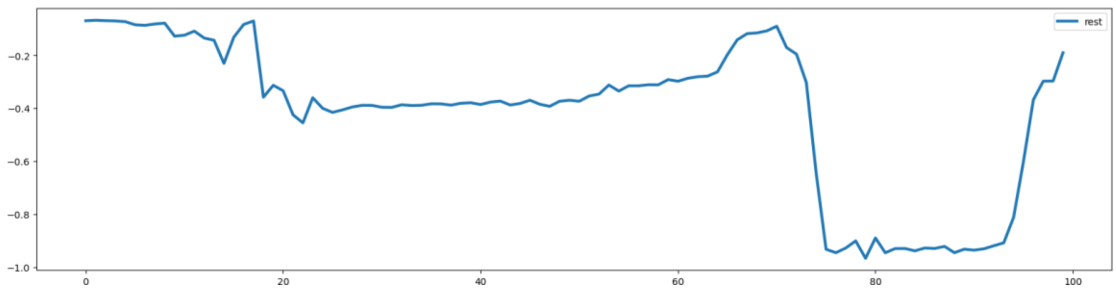






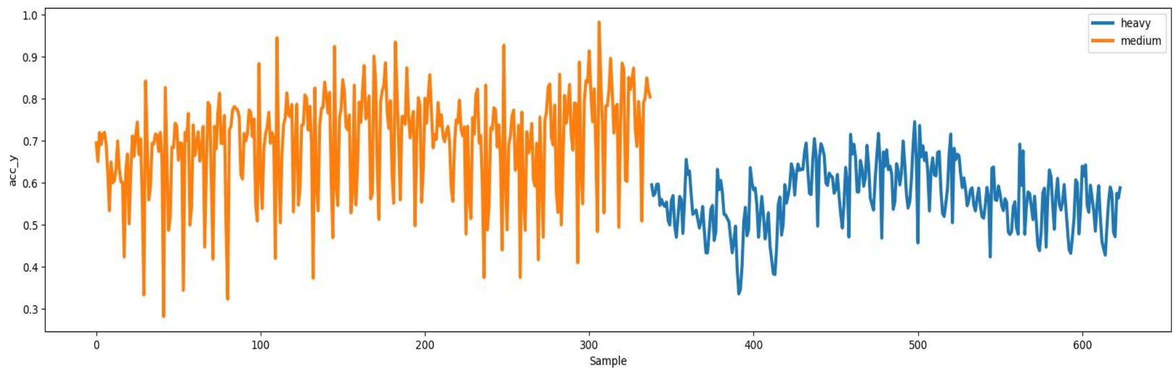






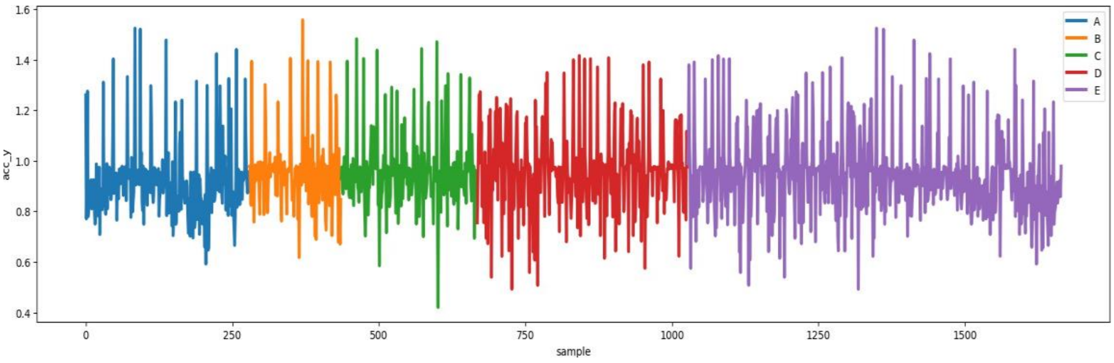
Category-wise Analysis (Squat):

- Conducted a category-wise analysis by grouping and plotting 'squat' exercise data.
- Examined variations under different conditions.



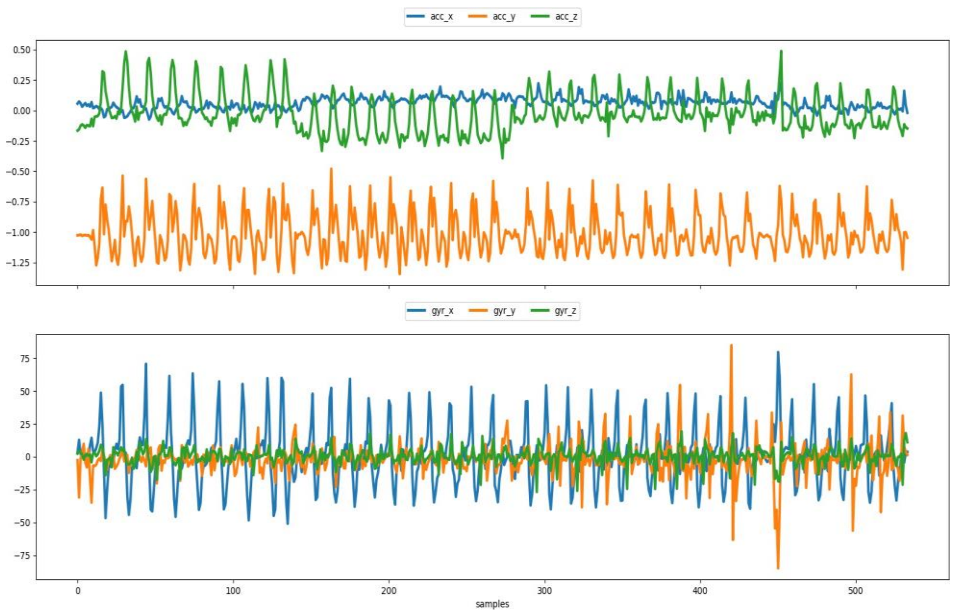
Participant-wise Analysis (Bench):

- Investigated 'bench' exercise data, considering different participants.
- Grouped by participant for a nuanced understanding.

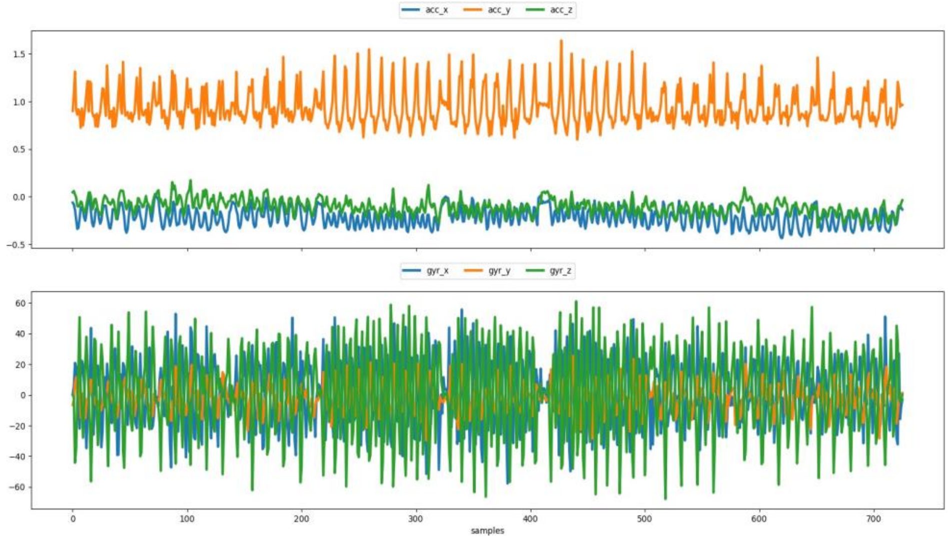


Combined Accelerometer & Gyroscope Plots:

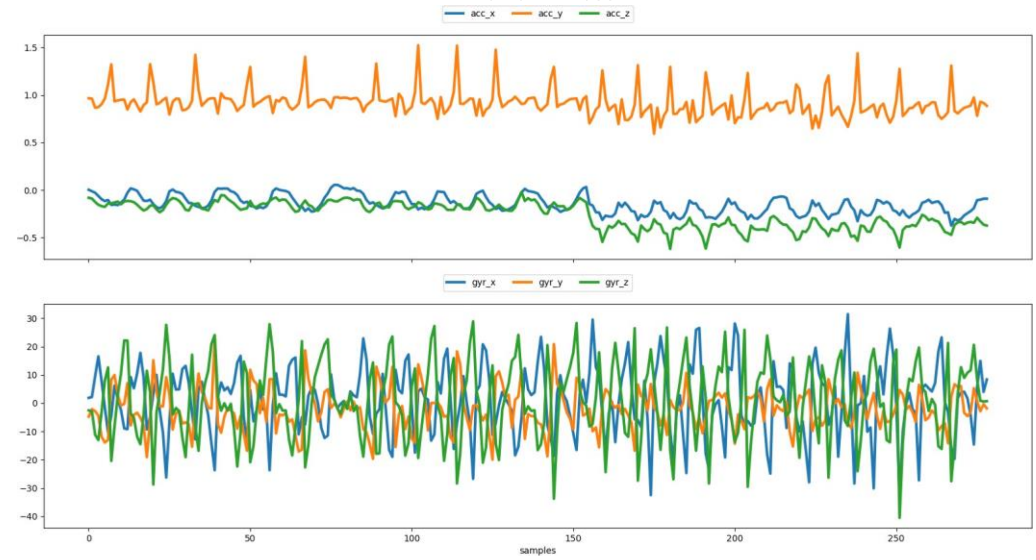
- Synthesized information by creating combined plots for both accelerometer and gyroscope data.
- Enabled a holistic view of sensor readings.
- For Analyse the Pattern we take for Visualization Participant A



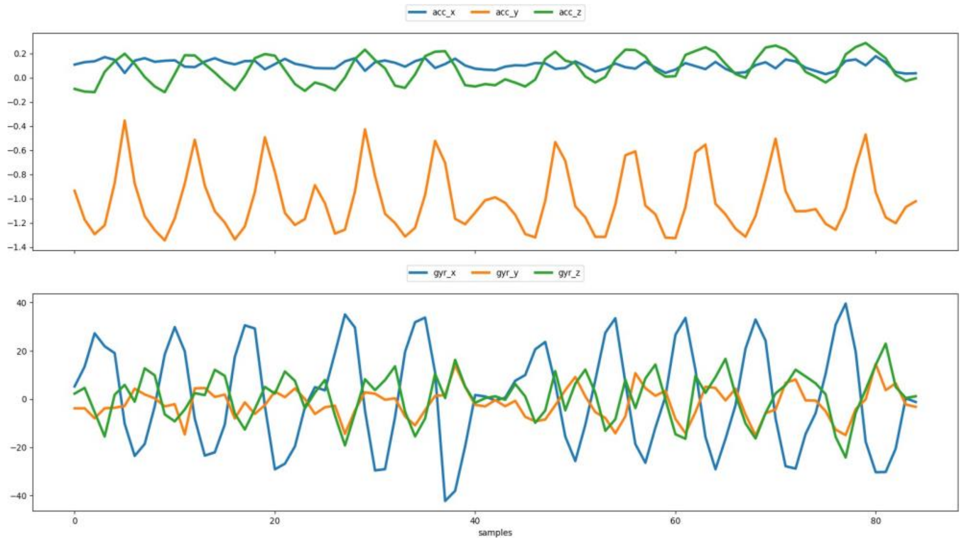
Overhead Press



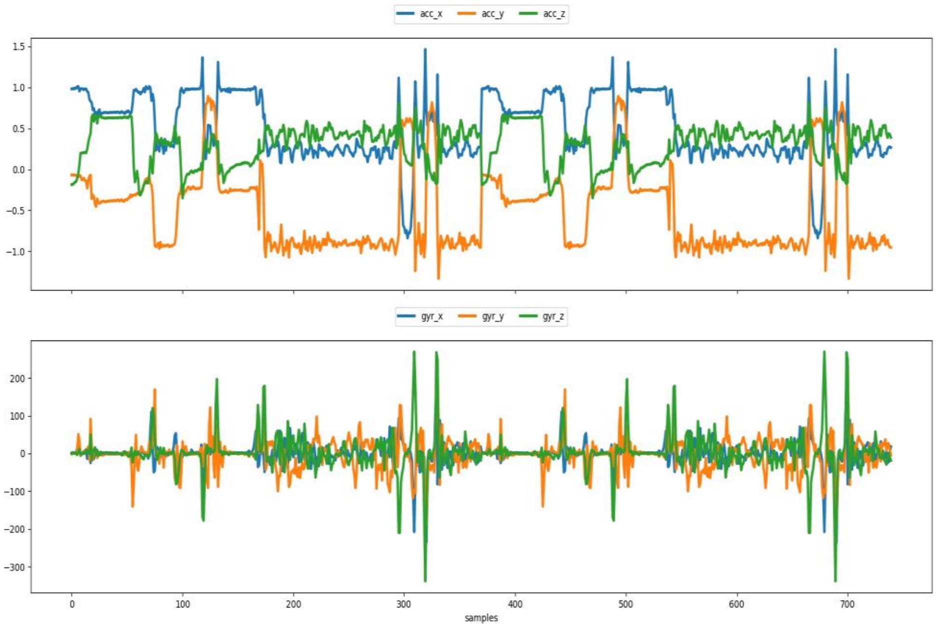
Bench Press



ROW


acc_x acc_y acc_z
gyr_x gyr_y gyr_z
samples

REST


acc_x acc_y acc_z
gyr_x gyr_y gyr_z
samples

HANDLING OUTLIERS

Subset Column :

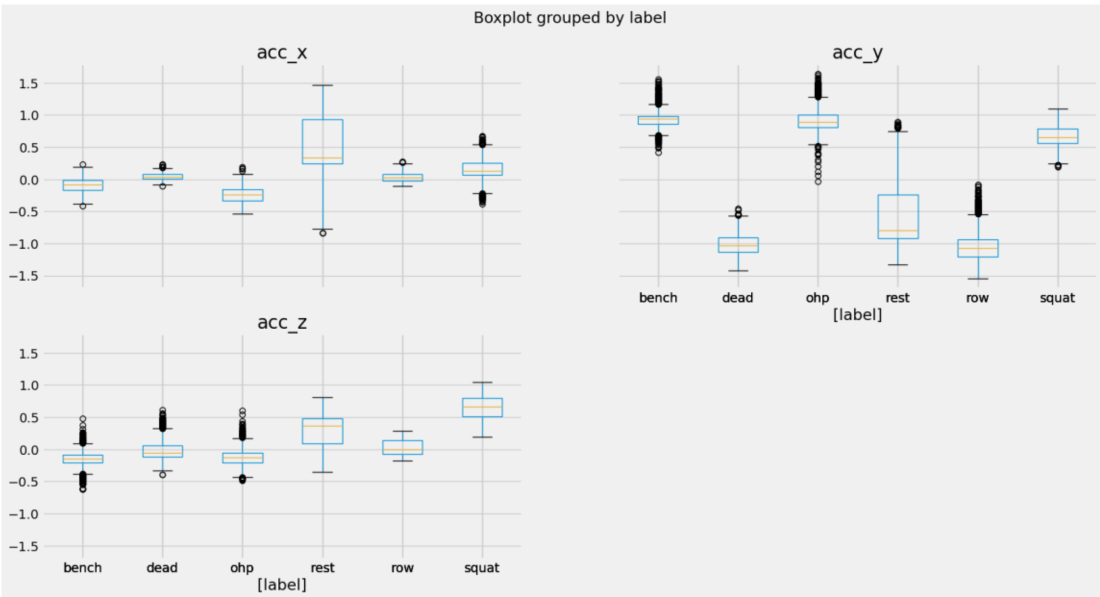
- ['acc_x', 'acc_y', 'acc_z', 'gyr_x', 'gyr_y', 'gyr_z']

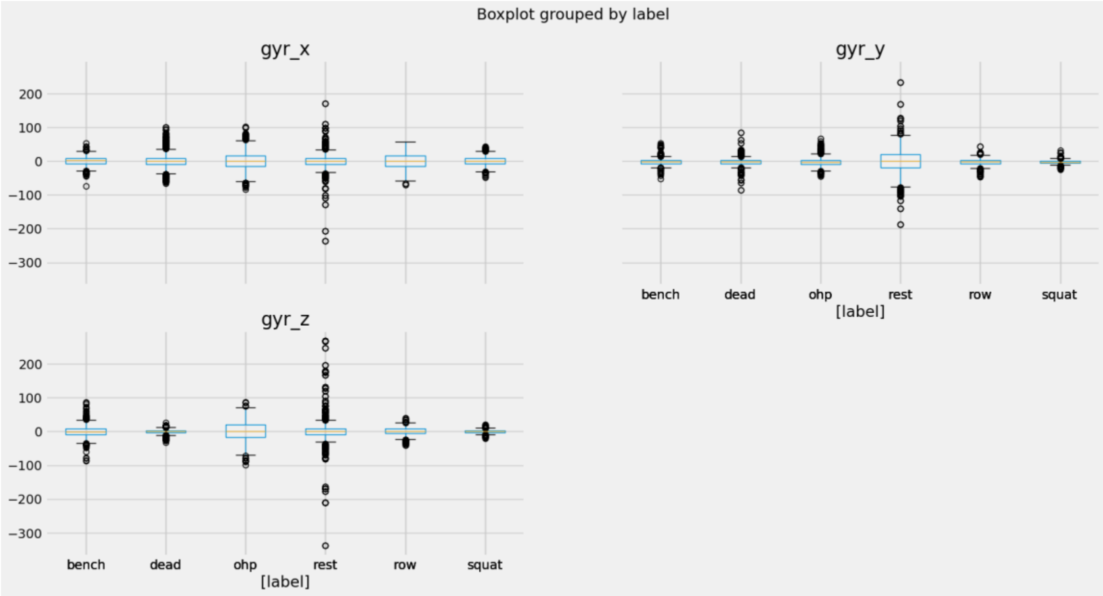
Outlier Detection Techniques:

- Utilized various techniques, including IQR, Chauvenet's Criterion, and Local Outlier Factor (LOF) on subset Visualized outliers through box plots and histograms.

Box Plots for outlier columns:

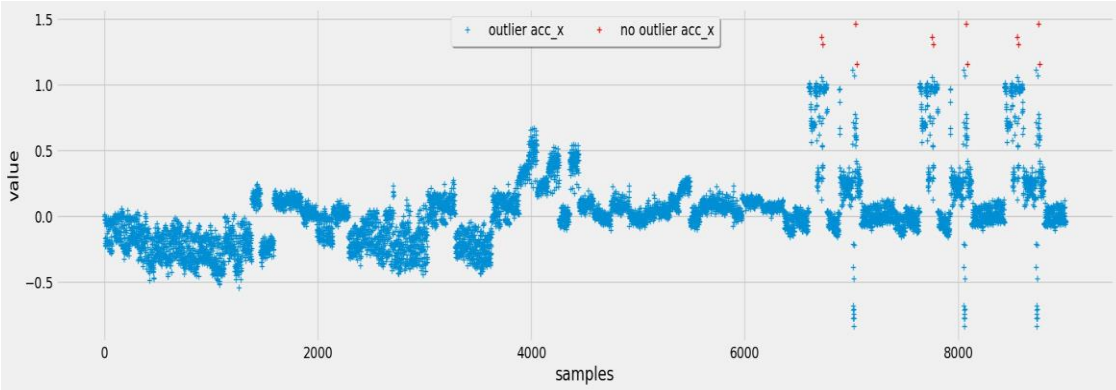
- Explored outliers in subset columns using box plots. Identified variations in different exercise labels.





Chauvenet's Criterion Visualization:

- Employed Chauvenet's Criterion to mark outliers.
- Visualized outlier points with binary markers on the original data



Handling Outliers - Opting for Chauvenet's Criterion:

- Chose Chauvenet's Criterion over other methods due to its superior performance, especially noticeable during visualization.

FEATURE ENGINEERING

Imputation of Missing Values:

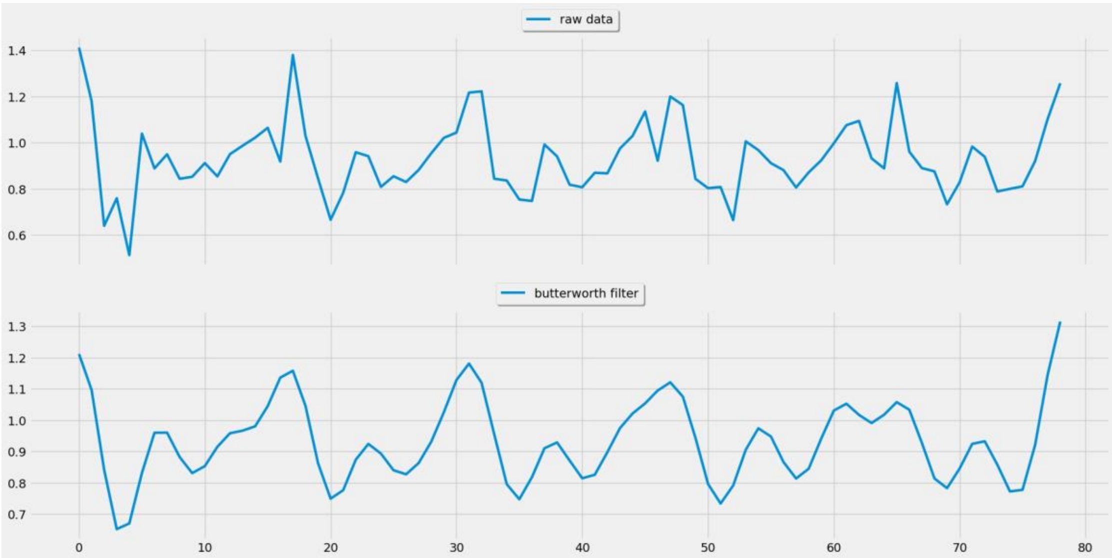
- Conducted interpolation for missing values in predictor columns to ensure data completeness.

Set Duration Calculation:

- Computed the duration of each set, allowing for temporal analysis.

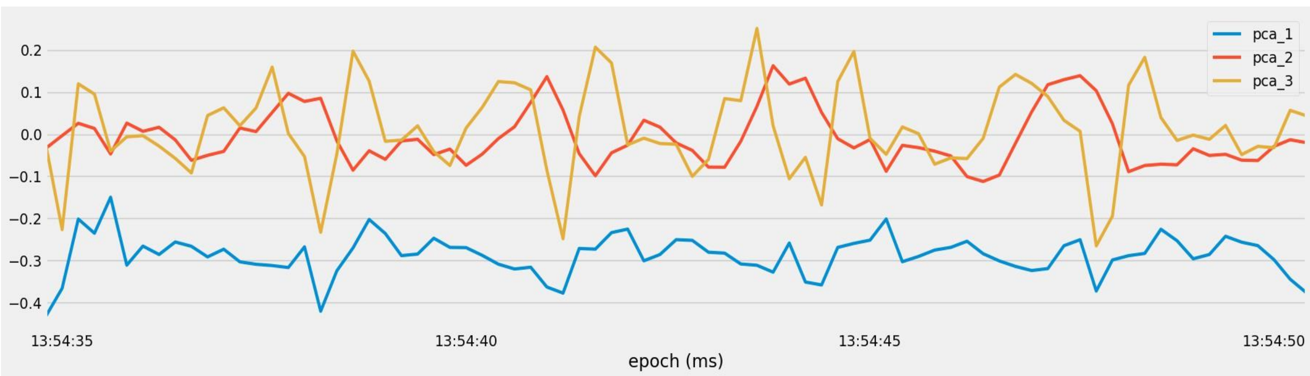
Butterworth Lowpass Filter:

- Applied a Butterworth lowpass filter to attenuate high-frequency noise, enhancing the quality of subset columns.



Principal Component Analysis (PCA):

- Utilized PCA to reduce dimensionality, preserving essential information, and visualized the top three principal components for set 45.



Temporal Abstraction:

- Employed temporal abstraction with a window size of 5 seconds to derive mean and standard deviation features for both accelerometer and gyroscope data.

Frequency Features:

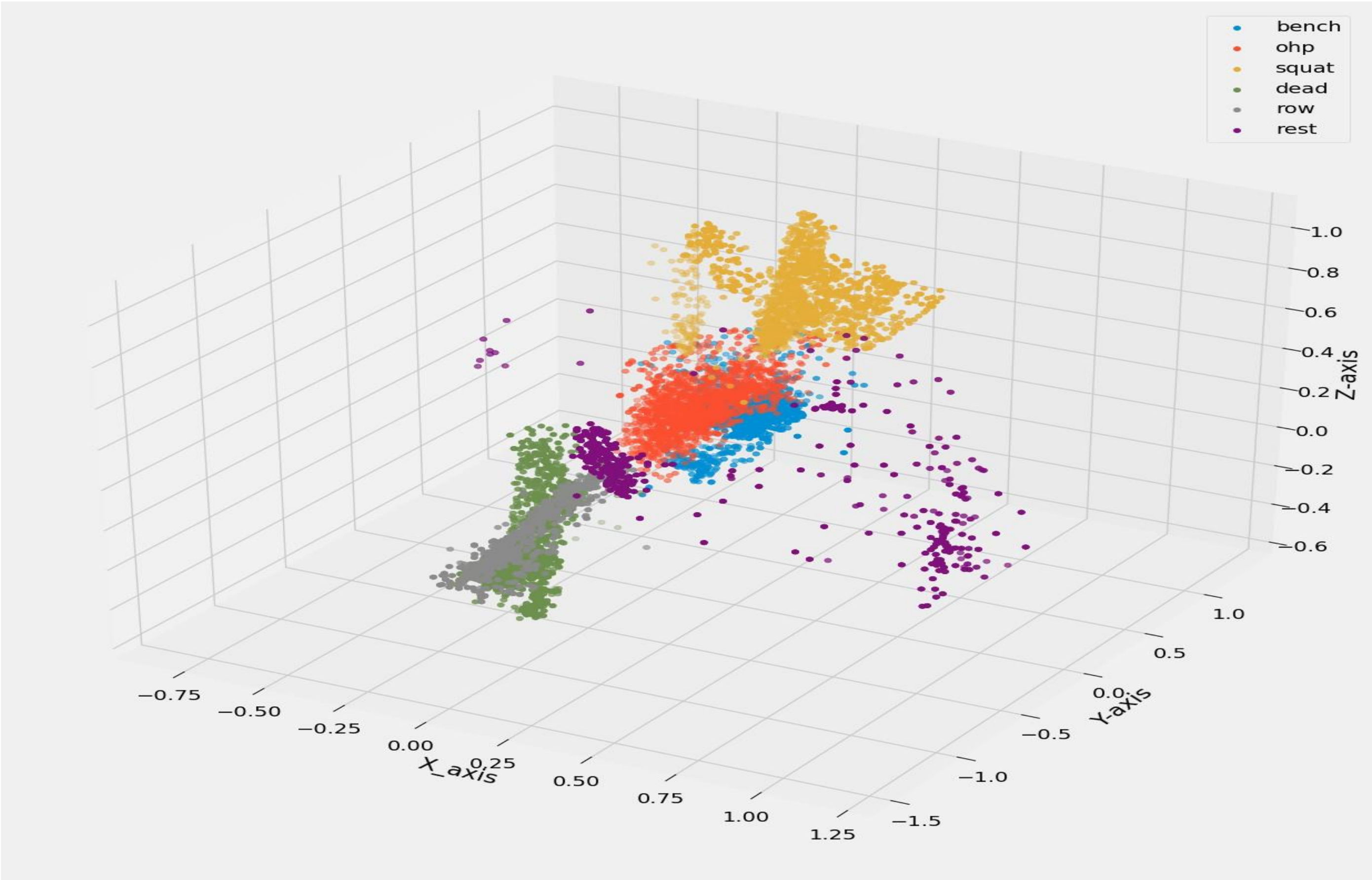
- Extracted frequency-related features, such as maximum frequency, weighted frequency, power spectral entropy, and specific frequency bands.

Overlapping Windows Handling:

- Addressed issues with overlapping windows, ensuring data consistency and reducing redundancy.

Clustering Analysis:

- Employed KMeans clustering with varying 'k' values and visualized clusters in a 3D plot based on accelerometer data.



TRAIN & TEST MODEL

Data Splitting:

- Created training and test sets from feature-engineered data.
- Visualized label distribution for total, training, and test sets.

Feature Sets:

- Defined multiple feature sets, including basic, square, PCA, temporal, frequency, and clustering features.
- Utilized forward feature selection to identify the top features for classification.

Model Comparison:

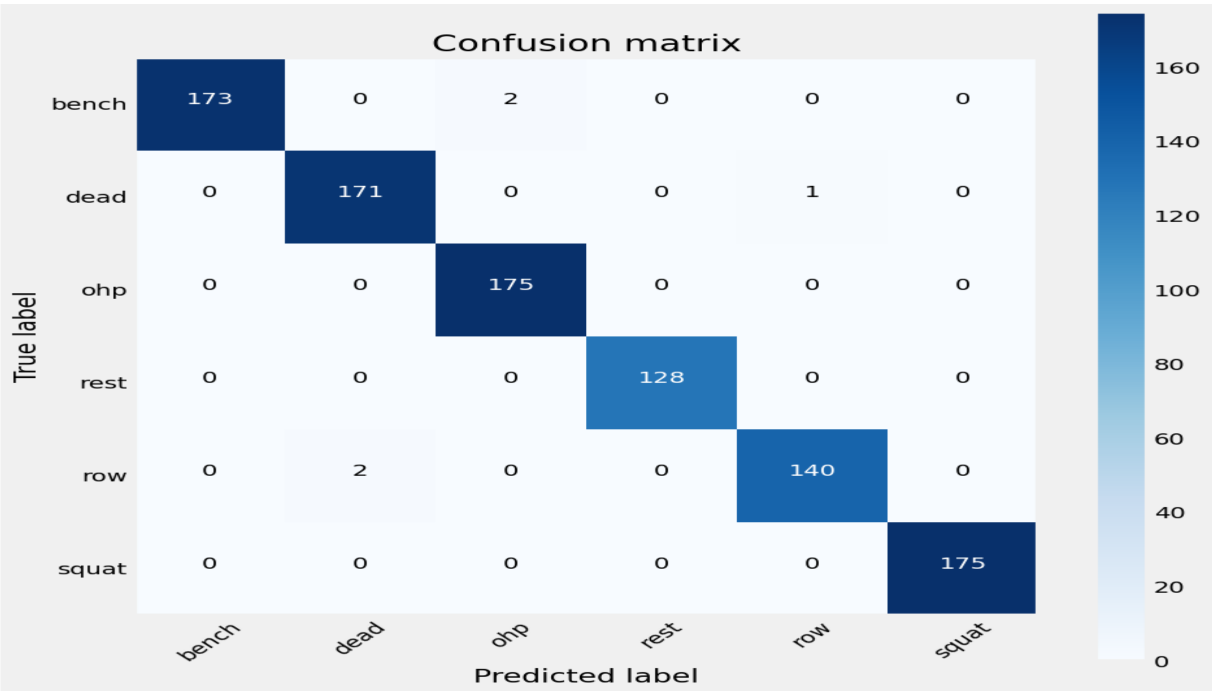
- Explored the performance of various classification models:
- Neural Network (NN)
- Random Forest (RF)
- K-Nearest Neighbors (KNN)
- Decision Tree (DT)
- Naive Bayes (NB)
- Conducted multiple iterations to account for non-deterministic classifiers.

Feature Set Impact:

- Investigated the influence of different feature sets on model accuracy.
- Analyzed accuracy scores for each model across various feature sets.

Best Model Evaluation:

- Selected Random Forest as the best-performing model based on feature set 4 with accuracy of 0.99.
- Evaluated accuracy and presented a confusion matrix for better understanding.



RESULTS

- The accuracy of RF is too good is about 0.99 and the least suitable model is KNN with accuracy of 0.78.
- This Model is Capable to predict the label of the exercise very effectively by taking input data of participant.

	Model	FeatureSet	Accuracy
1	RF	Feature Set 4	0.993795
0	NN	Feature Set 4	0.993795
1	RF	Selected Feature	0.991727
0	NN	Feature Set 3	0.991727
1	RF	Feature Set 3	0.985522
3	DT	Feature Set 4	0.984488
3	DT	Selected Feature	0.981386
0	NN	Selected Feature	0.972079
2	KNN	Feature Set 4	0.967942
1	RF	Feature Set 1	0.964840
4	NB	Feature Set 4	0.963806
3	DT	Feature Set 3	0.960703
1	RF	Feature Set 2	0.958635
4	NB	Feature Set 3	0.954498
0	NN	Feature Set 1	0.937952
0	NN	Feature Set 2	0.936918
2	KNN	Feature Set 3	0.931748
3	DT	Feature Set 2	0.930714
3	DT	Feature Set 1	0.927611
4	NB	Selected Feature	0.915202
4	NB	Feature Set 2	0.888314
4	NB	Feature Set 1	0.876939
2	KNN	Selected Feature	0.830403
2	KNN	Feature Set 1	0.812823
2	KNN	Feature Set 2	0.7983

CONCLUSION

This project delves into Gyroscope (gyro) and Accelerometer (accelo) data for human activity recognition, employing extensive data preprocessing and advanced machine learning techniques. From feature engineering to model development, various algorithms were explored, including Decision Trees, Random Forests, K-Nearest Neighbors, and Neural Networks. The evaluation process, utilizing accuracy assessments and confusion matrices, revealed insights into model performance across diverse feature sets. Beyond academia, this work has practical applications in real-world human activity recognition systems, with implications for healthcare, sports science, and assistive technology. The project contributes to the evolving landscape of precision and efficiency in activity recognition, aligning with the advancements in technology.

REFERENCE

https://www.youtube.com/watch?v=cCONIdrM2VI&list=PL-Y17yukoyy0sT2hoSQxn1TdV0J7-MX4K&ab_channel=DaveEbbelaar

https://mbientlab.com/metamotions/

https://github.com/mhoogen/ML4QS/

https://docs.datalumina.io/jD1BSJCAPYKSwh/b/D14B12A4-F691-4004-B381-692865344A5D/Chauvenet%E2%80%99s-Criterion