Final Summary of the UCI HAR Notebook

1. Objective

The goal of this project is to classify human activities using the UCI Human Activity Recognition (HAR) Dataset. The study explores two different modeling approaches:

1. Deep Learning Approach
   * Trains LSTM (Long Short-Term Memory) models on raw sensor data (accelerometer and gyroscope readings).
   * Does not use precomputed features provided by the dataset authors.
2. Machine Learning Approach
   * Extracts features from raw sensor data using TSFEL (Time-Series Feature Extraction Library).
   * Trains models such as Random Forest, SVM (Support Vector Machine), and Logistic Regression on these extracted features.

The notebook aims to compare the performance of these approaches and analyze which method is more effective for activity classification.

2. Dataset Used

The UCI HAR Dataset consists of smartphone sensor readings collected from 30 participants performing six different activities:

* WALKING
* WALKING\_UPSTAIRS
* WALKING\_DOWNSTAIRS
* SITTING
* STANDING
* LAYING

The dataset contains raw accelerometer and gyroscope readings along three axes (X, Y, and Z), recorded at 50Hz frequency.

For this study:

* Deep Learning models were trained directly on raw time-series sensor data.
* Machine Learning models were trained on features extracted from raw data using TSFEL.
* Precomputed features provided by the dataset authors were not used.

3. Models and Their Performance

The following models were trained and evaluated:

| Model | Data Used | Accuracy (%) |
| --- | --- | --- |
| Random Forest | TSFEL-Extracted Features | 67 |
| SVM (Support Vector Machine) | TSFEL-Extracted Features | 62 |
| LSTM (Long Short-Term Memory) | Raw Sensor Data | 61 |
| Logistic Regression | TSFEL-Extracted Features | 50 |

4. Key Observations and Comparisons

1. Machine Learning Models vs. Deep Learning Models
   * Machine Learning models trained on TSFEL-extracted features outperformed Deep Learning models trained on raw data.
   * Random Forest achieved the highest accuracy (67%), surpassing both SVM (62%) and LSTM (61%).
   * LSTM, the best Deep Learning model, could not outperform Random Forest or SVM.
2. Performance of Individual Models
   * Random Forest performed the best (67%), showing that traditional feature-based learning works well for this dataset.
   * SVM achieved 62% accuracy, slightly behind Random Forest but still higher than LSTM.
   * LSTM (61%) was the best-performing Deep Learning model, but still lagged behind ML models.
   * Logistic Regression had the lowest accuracy (50%), indicating that simple linear models are not effective for this task.
3. Impact of Feature Engineering (TSFEL)
   * Extracting features using TSFEL significantly improved the performance of ML models.
   * Raw sensor data, when used directly with LSTMs, did not lead to better classification than TSFEL-based feature extraction.

5. Conclusion

* Machine Learning models trained on TSFEL features outperformed Deep Learning models trained on raw sensor data.
* Random Forest was the best-performing model with 67% accuracy.
* Deep Learning (LSTM) showed promising results but did not surpass ML models, indicating that raw sensor data alone may not be sufficient.
* Feature extraction using TSFEL played a crucial role in improving classification accuracy for ML models.
* Future improvements could explore combining deep learning with TSFEL-extracted features to enhance performance.