## Feature Engineering

* + Extraction of time-domain and frequency-domain features from sensor data.

# 🕒 Time Domain Features

* + These are calculated directly from the raw signal over time.

|  |  |
| --- | --- |
| Feature | Description |
| Mean | Average of the signal values. |
| Standard Deviation | Measures variability or dispersion. |
| Variance | The square of the standard deviation. |
| Min / Max | Minimum or maximum value in the time window. |
| Range | Difference between max and min values. |
| Median | The middle value of the signal. |
| Interquartile Range (IQR) | Difference between the 75th and 25th percentiles. |
| Root Mean Square (RMS) | Square root of the average of squared values. |
| Zero-Crossing Rate | Number of times the signal crosses zero. |
| Skewness | Measure of signal asymmetry. |
| Kurtosis | Measure of the 'tailedness' of the signal distribution. |
| Signal Magnitude Area (SMA) | Combined magnitude from 3-axes: (1/N)  ∑(|x| + |y| + |z|) |
| Autocorrelation | Similarity of a signal with a delayed version of itself. |
| Peak Count / Peak Amplitude | Number of peaks or amplitude of peaks in the window. |
| Energy | Sum of squared values: ∑(x²) |

# 📉 Frequency Domain Features

* + These are obtained by transforming the signal using methods like the Fast Fourier Transform (FFT).

|  |  |
| --- | --- |
| Feature | Description |
| Spectral Centroid | Indicates the center of mass of the spectrum. |
| Spectral Entropy | Measure of the signal complexity in frequency domain. |
| Spectral Energy | Sum of squared FFT components. |
| Dominant Frequency | Frequency with the highest magnitude. |
| Frequency Variance | Variance of the spectral components. |
| Spectral Flatness | Ratio of geometric mean to arithmetic mean; indicates tonality. |
| Peak Frequency | Frequency with maximum power. |
| Bandwidth | Range between the frequencies where most energy is concentrated. |
| FFT Coefficients | Raw coefficients (typically first few) from FFT used as features. |
| Power Spectral Density (PSD) | Energy distribution across frequency. |

* + Feature selection techniques to identify the most informative features.
    - Implement techniques from random forest-based models to reduce the feature set and improve model performance.

## Model Development

* + Explore and compare different machine learning models, including decision trees, KNN (K- Nearest Neighbors), support vector machines and random forests.

## Evaluation

* + Implement 10-fold cross-validation and Leave One Subject Out/ Leave One Episode Out cross-validation to evaluate model performance.
  + Use appropriate metrics (accuracy, F1 score, precision, recall) to evaluate model performance.
  + Discuss the performance of different models and feature sets.

## Conclusion

* + Summarize key findings and lessons learned.
  + Discuss potential improvements and future research directions.

**Project Deliverables:**

## Demo 3 Presentation (Present items 1–3 during your scheduled time)

* + Include clear and concise code demonstrations showing how to preprocess the dataset, train models, and evaluate their performance.

## Project Deliverables (Submit a hard copy of your report in person and upload your code to Canvas on May 6th or 8th)

* + **Report**: The report should detail every step, including all methodologies, experimental results, and insights. Please bring a hard copy on **May 6th or 8th** during class and submit it in person.

**Code**: Ensure the codebase is well-documented, with clear instructions on running the preprocessing, training, and evaluation scripts. Submit your code on Canvas by **May 6th** or 8th.