Important Notes

1. 30 second windows capture enough physiological changes to make accurate classifications
   1. Anything shorter or longer would mean the changes would not be detected accurately.
2. EDA, electrodermal activity, shows the electrical conductance of your skin which changes in moisture/sweat level within an individual’s skin
   1. Response to external stimuli
      1. Showcases stress level, arousal level during sleep
         1. Can be used to classify sleep stage
   2. Sleep Stage Indicators:
      1. **Wake/Light Sleep**: Higher EDA due to stress, anxiety, or environmental responses
      2. **Deep Sleep** (N3): Lower EDA as the body is more relaxed and less reactive
      3. **REM Sleep**: Variable EDA - can spike during intense dreams
      4. **Sleep Transitions**: EDA often changes when moving between sleep stages
3. The dataset introduces terms like **PPG** (photoplethysmography) sensor which can be found in digital watches that measures your heart rate and hr variability

The dataset has these columns:

* **Timestamp**
  + Tells the time throughout the night
* **BPV (Blood Volume Pulse)** 
  + How much blood is rushing through your wrist at each moment
    - The watch shines a light into your skin
    - More blood = absorbs more light = different reading
    - Less blood = absorbs less light = different reading
* **IBI (Inter-beat interval)** 
  + The time gap between two heartbeats
    - Awake/stressed: Heart rhythm is more irregular (variable IBIs)
    - Deep sleep: Heart rhythm is more steady (consistent IBIs)
    - REM sleep: Heart rhythm gets irregular again (variable IBIs)
* **EDA (Electrodermal activity)**
  + shows the electrical conductance of your skin which changes in moisture/sweat level within an individual’s skin
  + Response to external stimuli
    - Showcases stress level, arousal level during sleep
      * Can be used to classify sleep stage
  + Sleep Stage Indicators:
    - **Wake/Light Sleep**: Higher EDA due to stress, anxiety, or environmental responses
    - **Deep Sleep** (N3): Lower EDA as the body is more relaxed and less reactive
    - **REM Sleep**: Variable EDA - can spike during intense dreams
    - **Sleep Transitions**: EDA often changes when moving between sleep stages
* **TEMP**
  + Temperature of the skin
* **ACC\_X, Y,Z**
  + Movement of the wrist
* **HR**
  + Heart rate
* **Different sleep stages** 
  + P (Preparation) - Setting up equipment, not actually sleeping yet
  + W (Wake) - Fully awake and conscious
  + N1 (Light Sleep) - Just falling asleep, easily woken up, drowsy
  + N2 (Light Sleep) - Proper light sleep, harder to wake up than N1
  + N3 (Deep Sleep) - Very deep sleep, hard to wake up, most restorative
  + R (REM Sleep) - Dream sleep, brain very active, eyes moving rapidly

BASLINE

**XGBoost Quick Reference**

**What is XGBoost?**

* XGBoost = eXtreme Gradient Boosting
* Ensemble method that combines multiple decision trees
* Uses gradient boosting - trains trees sequentially to fix previous errors
* Best for structured/tabular data problems

**How XGBoost Works**

* Builds trees one at a time, each correcting errors of previous trees
* Uses gradient descent to minimize loss function + regularization
* Grows trees level by level using greedy splits
* Each split chosen based on gain score (improvement in objective)

**Key XGBoost Features**

* Built-in regularization (L1 and L2) to prevent overfitting
* Automatic handling of missing values
* Built-in cross-validation support
* Parallel processing for speed
* Feature importance calculation (gain, cover, frequency)

**When to Use XGBoost**

* Structured/tabular data with mixed feature types
* Need high predictive accuracy
* Want feature importance insights
* Have missing values in data
* Competition or production environments

**XGBoost Advantages**

* High performance on tabular data
* Handles missing values automatically
* Robust to outliers
* No need for feature scaling
* Fast training with parallel processing
* Built-in regularization prevents overfitting