PROJECT 5 Symptom-Onset prediction with LSTM and sliding window

Business scenario

Wearables capture continuous vitals; clinicians need to spot the *day* a symptom becomes acute.

Dataset

https://www.kaggle.com/datasets/nasirayub2/human-vital-sign-dataset

or synthetic/similar dataset

Core technique

A unidirectional or bidirectional **LSTM** that outputs a 3-class tag (acute / chronic / unknown) for each day in a 30-day window.

Key steps

- Create sliding 30-day windows; apply label smoothing for borderline days.

- Train the LSTM; compare to a logistic-regression baseline.

- Plot per-timestep predictions and ground truth.

Deliverables

- Notebook with exploratory data analysis (EDA).

- Training curves and confusion matrices.

- Brief write-up of error patterns (e.g., weekend gaps).

- **Per-timestep F1** and **AUROC** on the test split.

- Comparative table: LSTM vs. baseline.

- Discussion of false-positive clusters.