SAIL 2025 Python Project Proposal

Group 23

Team Member: Junlin Li, Yuxuan Li, Kristof Hollstein, Ida Bagus Rai Satria Dharma, Muhammad Harisuddin

1. Objectives

2. Research Questions

3. Intended Data Analysis

4. List of Datasets

5. Geographical Scale

Proposal: Predictive Dashboard for Crowd Flows at SAIL Amsterdam 2025

1. Objectives

- Develop a dashboard prototype to **monitor and predict short-term crowd flows** during SAIL Amsterdam 2025.
- Provide the present and **forecasts** of crowd densities at specific areas.
- Enable event managers to **anticipate congestion and safety risks** by visualizing real-time data, predictions, and anomalies on an **interactive map interface**.
- To serve as a reference material to prevent undesirable incidents in the future

2. Research Questions

1. How accurately can crowd flow patterns be predicted in the short term (15–120 minutes) using only sensor data?
2. What simple forecasting methods (e.g., moving averages, autoregression, sliding window retraining) are effective in reducing bias?
3. How can the dashboard best communicate status, forecasts, and anomalies to operators for quick decision-making?
4. What is the added value of using an **interactive map visualization with hover tooltips** compared to static charts?

3. Intended Data Analysis

- **Data Preprocessing**:
  - Clean missing values, align time intervals (3-minute resolution).
  - Aggregate and smooth data to reveal trends.
- **Exploratory Analysis**:
  - Identify temporal patterns (daily peaks, event-related surges).
  - Detect variability across locations.
- **Forecasting Approaches**:
  - Baseline models: Moving Average, Autoregressive (AR).
  - Online correction: Trend correction, sliding window retraining.

    - Uncertainty representation: Mean ± standard deviation to generate forecast intervals.
- **Evaluation Metrics**:
    - RMSE/MAE for overall prediction accuracy.
    - Peak timing error (difference between predicted vs. observed peak).
    - Anomaly count (flagging unrealistic predictions, e.g., sudden extreme values).

4. List of Datasets

- **Confirmed dataset**:
    - *Crowd flows from SAIL sensors* (3-minute intervals).
- **Potential extensions (not guaranteed, but optional later)**:
    - KNMI weather data (temperature, precipitation, wind).
    - Traffic flow data
    - AIS ship location data.

For this proposal, the focus remains on **crowd flows only**.

5. Geographical Scale

- **Study area**: Key crowd gathering zones in Amsterdam during SAIL Amsterdam Area
- **Spatial resolution**: Individual sensor locations as points on the map.
- **Visualization**:
    - Each sensor represented as a point colored by density.
    - Hover tooltips show detailed info (current flow, forecast, anomaly status).
    - Switchable layers for different forecast horizons (current, +15 min, +1h, +2h).