Proposal – Short-term Crowd Predictions for SAIL 2025

Authors:
Wessel Toutenhoofd 4967402
Mats Jager - 5387906
Abel Arts - 5516064

1. Introduction & Motivation

The SAIL event in Amsterdam is the largest nautical event in the world, attracting millions of visitors over several days. Ensuring safe and efficient crowd movement is a critical challenge in such large-scale urban events.
Our project aims to build a short-term prediction system for crowd density and flows in the IJhaven and surrounding event areas to prevent overcrowding. This will serve as an early-warning tool to prevent congestion, improve safety, and optimise crowd management decisions.

2. Research Objectives & Questions

Objective:
To design and evaluate a predictive model that can forecast crowd densities and movement patterns 5–15 minutes ahead, using crowd sensor data

Research Question:

> *To what extent can real-time and contextual data be used to accurately predict short-term (5–15 minute) crowd behaviour patterns during SAIL 2025?*

Specific: The study focuses on the SAIL 2025 event, making use of collected real-time and contextual data to model and predict crowd behaviour, focusing on one specific location (Amsterdam) and one time period (20-24 of August 2025).

Measurable: Predictions will be compared against real-world observations to accurately assess model performance.

Achievable: SAIL is a well-documented, large-scale event with rich data availability, making it an ideal case study for developing and validating short-term crowd prediction models.

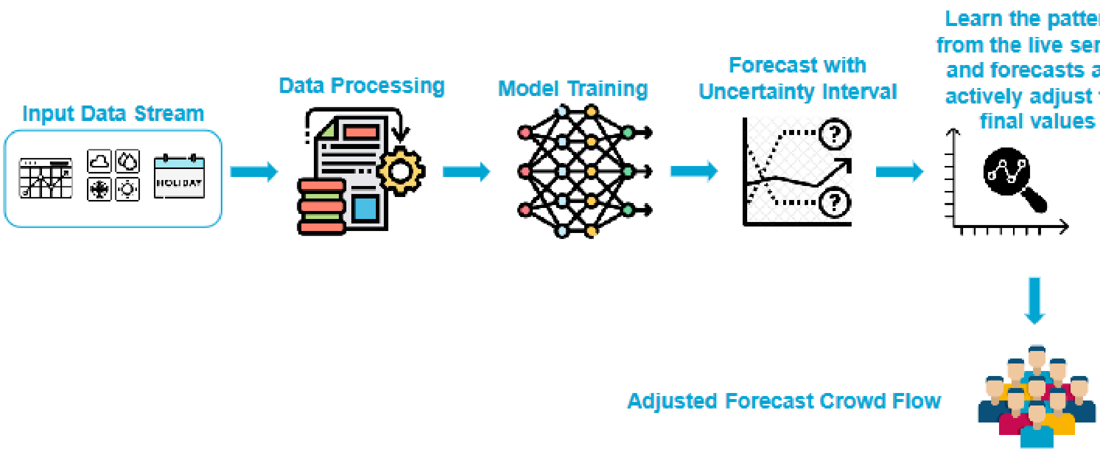
Relevant: SAIL is one of the largest global events. Accurate short-term predictions contribute directly to operational decision-making and risk prevention.

Time-Bound: The study is limited to short-term prediction (5–15 minutes).

Subquestions

1. How can crowd surges be predicted in real time?
2. What methods can be used to quantify uncertainty in short-term predictions?
3. Which input features (public transport arrivals, time-of-day, weather, event schedule) contribute most to prediction accuracy?
4. How can these predictions be effectively visualised for use in crowd management dashboards?
5. How can AI be used to monitor crowd movements, predict bottlenecks and risks, and provide decision support?

3. Intended Data Analysis Pipeline



Source: TU Delft, Faculty of MSc Transport, Infrastructure and Logistics, TIL Research and Design Methods lecture slides

In the figure, an example of a data analysis pipeline is visualised for large events like SAIL. This is dissolved in the following steps:

1. **Input data stream**
    - Collect raw data from available sources (see datasets below). In real-time events, this involves the data obtained from crowd sensors.
    - Weather data.
    - Temporal data.
    - Geospatial alignment: map crowd densities to specific event zones.
    - Use of AI to monitor crowd movement.

2. **Data Processing and Model Training**
    - Transformation of raw data into usable information.
    - Descriptive statistics of crowd density and flows.
    - Visual exploration: understanding the data using graphical methods.

3. **Prediction Models**

    - Forecasting methods to predict surges and other worrying flows.
    - The use of AI to predict bottlenecks and risks.

4. **Visualization**
    - Heatmaps of predicted crowd densities.
    - Time-series plots of actual vs. predicted flows.

5. **Evaluation**
    - Learn the patterns and actively adjust the final values.
    - Verify insights by cross-checking them with the on-duty traffic controllers.

For data monitoring in short-term predictions, one should be aware that the system should be operating as a feedback system; The situation is constantly changing, which means the output should be used in the input of the design, to be efficient at constant times.

4. Dataset(s)

We will combine **event-specific SAIL 2025 data** with open datasets regarding crowd control to enrich predictions:

SAIL 2025 Event Data

- **Crowd sensors:** In addition to camera counts, social media monitoring, and timestamp counts, Wi-Fi sensors and automatic counting gates are also included. Research shows that data fusion of Wi-Fi traces with automatic counting systems improves the estimation of pedestrian flows.

- **Event schedules:** Start and end times of parades, fireworks, music performances, ship visits, etc. This information helps to explain peaks in crowd density.

- **Spatial layout:** Map of the IJ harbour with zones, entrances, bridges, bottlenecks, and emergency exits.

- **Logistical metadata:** Information on maximum visitor capacity, queue lengths, temporary closures, and crowd-control measures (e.g., dynamic gates).

Amsterdam Public Transport Data

- **GVB/NS arrivals:** Arrival and departure data for trains, trams, buses, and ferries at stops around the SAIL area (NDOV Loket / data.overheid.nl).

- **Vehicle occupancy:** Where available, occupancy rates per trip (for example, through OV-chipkaart statistics).

- **Bicycles and shared mobility:** Numbers of OV-bikes, shared bikes, and e-scooters picked up or returned near SAIL.

Road Traffic Data (NDW)

- **Traffic intensity and speed:** Number of vehicles and average speed per measurement point

- **Situational data:** Roadworks, temporary traffic measures, traffic reports, safety messages (slippery roads, obstacles, etc.)

- **Bridge openings and lane status:** Notifications of bridge openings and (rush-hour) lane status

Aggregated Mobile Phone Data (CBS)

- **Mobile telephony metadata:** Anonymised, aggregated data from mobile phones is used by CBS to estimate the number of residents and visitors present in a municipality at a given time

Parking and Mobility

- **Parking spaces and occupancy:** data.amsterdam.nl provides open data on the number of parking spaces and the real-time occupancy of parking garages
  This helps predict car-related inflow.

- **P+R locations:** Occupancy of park-and-ride facilities.

- **Taxi and ride-hailing:** Number of taxis and rides departing from or arriving at the event area (if available via open data).

Smart Waste Management Sensors

- **Garbage/waste sensors:** In 2014, Amsterdam introduced waste-truck load monitoring and installed about 12,500 fill-level sensors in trash bins
  hologram.io. A sudden high fill rate may indicate increased crowd density.

Wi-Fi Sensors and Counting Gates

- **Wi-Fi traces and automatic counting gates:** Crowd monitoring systems combine Wi-Fi sensors with automatic counting systems to accurately estimate pedestrian flows *openresearch.amsterdam*. This is an important complement to existing crowd density measurements.

DrukteRadar / Crowd Flow Dashboard Sources

- The **Crowd Flow Dashboard working group** and **Amsterdam's DrukteRadar** combine datasets from public and private transport, parking facilities, hotel bookings, economic activity, tourism, scientific studies, waste management, mobile operators, and footfall analytics. *northsearegion.eu*.

- DrukteRadar and DrukteKalender use various data streams to calculate, visualise, and predict local crowding; the results are publicly shared *northsearegion.eu*. This approach inspires us to also explore tourist statistics, hotel occupancy, and spending patterns in addition to the sources mentioned above.

Other Contextual Data

- **Weather data (KNMI Open Data):** Hourly values of precipitation, temperature, wind, and weather forecasts.

- **Google Mobility Reports:** Historical context of movement patterns at the neighbourhood level.

- **CBS Open Data:** Population density, age structure, and socio-economic indicators for Amsterdam districts.

- **Pedestrian/social media analyses:** Real-time sentiment or incident reports via hashtags or app feedback (with attention to privacy and representativeness).

5. Geographical and Temporal Scale

- Geographical Scale: IJhaven harbour and adjacent areas of Amsterdam (SAIL event zone).

- Temporal Scale: August 20–24, 2025 (SAIL event period). Predictions focus on short-term windows of 5–15 minutes.

6. Expected Outcomes

Our project aims to deliver both technical and practical outcomes that are useful for crowd management during SAIL 2025. The expected outcomes are:

Functional Prediction Pipeline

A reproducible Python-based pipeline that takes raw event and contextual data as input and produces short-term crowd density forecasts. The pipeline will include modules for data preprocessing, feature extraction, model training, and real-time prediction.

Comparative Model Evaluation

A comparison of modelling approaches, including statistical baselines (ARIMA, exponential smoothing) and machine learning models (Random Forest, Gradient Boosting). We will evaluate them using metrics such as RMSE and MAE, while highlighting trade-offs between accuracy, complexity, and computational efficiency.

Insights into Influential Factors

By analysing feature importance, we aim to identify the main drivers of crowd fluctuations, such as PT arrivals, time-of-day, weather conditions, and event schedule. These insights can guide preventive strategies (e.g., staff reallocation, signage updates).

Visualisation and Dashboard Prototype

An interactive visualisation tool, likely in Streamlit, that displays predictions in a user-friendly way. Features will include:

- Real-time heatmaps of crowd density in the IJhaven and adjacent areas.
- Time-series plots comparing predicted vs. actual crowd densities.
- Alert indicators when thresholds of crowd density are likely to be exceeded.
- Interactive filters (time horizon, location, model type) for scenario exploration.

Contribution to Crowd Management Research

The outcomes will serve as a case study in applying short-term prediction models to large-scale events. The approach could also be extended to other festivals, sports events, or public gatherings.

Documentation and Reproducibility

A well-documented GitHub repository with reproducible notebooks, clear instructions, and dataset references will be delivered. This ensures transparency and enables future groups or researchers to build on our work.

7. References

- Daamen, W., Duives, D., Yuan, Y., & Hoogendoorn, S. (2018). *Real-time crowd monitoring dashboard*. TRB 2018 Annual Meeting.
- Gong, V. X., Daamen, W., Bozzon, A., & Hoogendoorn, S. P. (2020). Crowd characterization for crowd management using social media data in city events. Travel Behaviour and Society, 20, 192–212. https://doi.org/10.1016/j.tbs.2020.03.011
- Daamen, W. (n.d.). *TIL Research and Design Methods, SAIL2025 design challenges* (lecture slide 23). TU Delft, Faculty of MSc Transport, Infrastructure and Logistics.

8. GitHub & Collaboration Plan

- Repository: Public GitHub repo (group-shared).
- Workflow:
    - Each member works on a feature branch.
    - Weekly merge into main after peer review.
    - Clear commit messages + pull requests (to avoid merge conflicts)