Predicting Pedestrian Counts using Machine Learning

Molly Asher[1], Yannick Oswald[2], Nick Malleson[2]

[1] School of Earth and Environment, University of Leeds
[2] School of Geography, University of Leeds

These slides: https://urban-analytics.github.io/dust/presentations.html

UNIVERSITY OF LEEDS

Predicting Pedestrian Counts using Machine Learning

11:00-12:30
Theatre Hall

SESSIONS

S2a, Accessibility & trips
Moderator: Edward Verbree

Eva Nuhn, Kai Hamburger and Sabine Timpf, Urban sound mapping for wayfinding – A theoretical approach and an empirical study [1]

Molly Asher, Yannick Oswald, and Nick Malleson, Predicting Pedestrian Counts using Machine Learning

Hoda Allahbakhshi, Joris Senn, Nicola Maiani and Alexandra Georgescu, Spatial Accessibility Assessment of Homecare Workers to the Older Population in the City of Zurich

Nir Fulman, Maria Marinov, and Itzhak Benenson, Exploring Non-Routine Trips Through Smartcard Transaction Analysis

Manuscript

B Urban Analytics and City Science

EPB: Urban Analytics and City Science
2025, Vol. 0(0) 1–24

DOI: 10.1177/23998083251319058
journals.sagepub.com/home/epb
Sage

Understanding pedestrian dynamics using machine learning with real-time urban sensors

Molly Asher
University of Leeds, UK

Yannick Oswald
University of Lausanne, Switzerland

Nick Malleson
University of Leeds, UK

Abstract
Quantifying, understanding and predicting the number of pedestrians that are likely be present in a particular place and time ('footfall') is critical for many academic, business and policy questions. However, limited data availability and complexities in the behaviour of the underlying pedestrian 'system' make it extremely difficult to accurately model footfall. This paper presents a machine learning model that is trained on a combination of hourly footfall count data from sensors across a city as well as important contextual factors that are associated with pedestrian movements such as the structure of the built environment and local weather conditions. The aims are to better understand the relationship between various contextual factors and footfall and to predict footfall volumes across a spatially heterogeneous city. The case study area is the city of Melbourne, Australia, where abundant pedestrian count data exist. Time-related variables, particularly time-of-day and day-of-week, emerged as the most significant predictors. While some built environment factors such as the presence of certain landmarks and weather conditions were influential, they were less so than temporal cycles. Interestingly the model over-estimates footfall in the years following the COVID-19 pandemic, suggesting that urban dynamics have yet to return to pre-

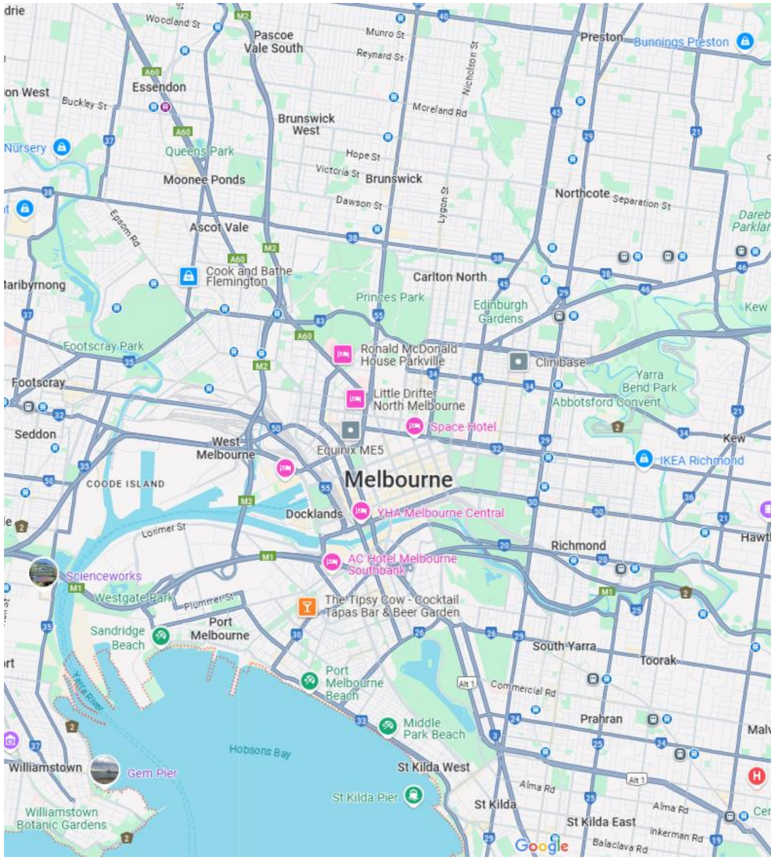
Modelling Overview

- Aim:
    - Use sensor data to build a predictive model which can estimate the number of people who will be at any location in the city at any given time

- To train a model, we needed a city with both:
    - Data on what we wanted to predict (dependent variable)
        - **number of pedestrians at different locations over time**
    - Data on which to base the predictions (explanatory variables):
        - **time (hour, day, month, year)**
        - **weather conditions**
        - **local built environment**
        - **connectedness of location (road betweenness)**

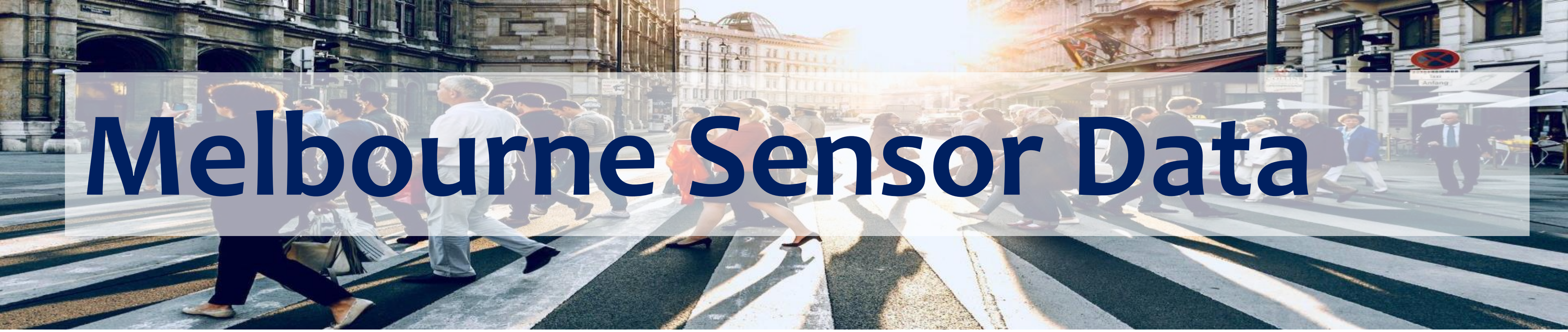
UNIVERSITY OF LEEDS

Melbourne Sensor Data

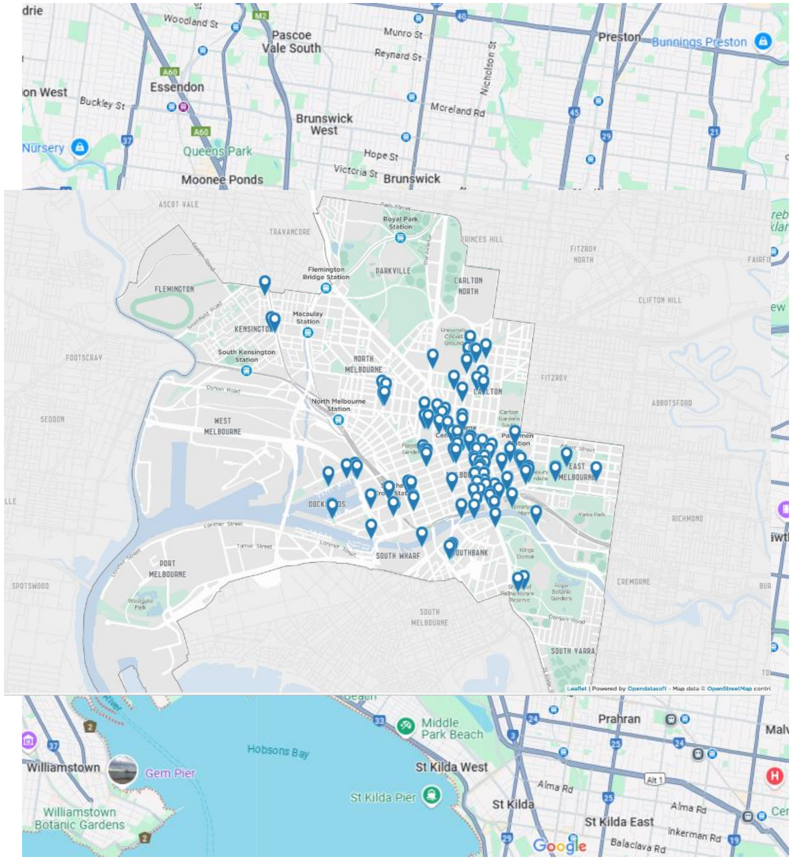


1 https://www.data.vic.gov.au/discover-city-melbourne-open-data

UNIVERSITY OF LEEDS



Melbourne Sensor Data

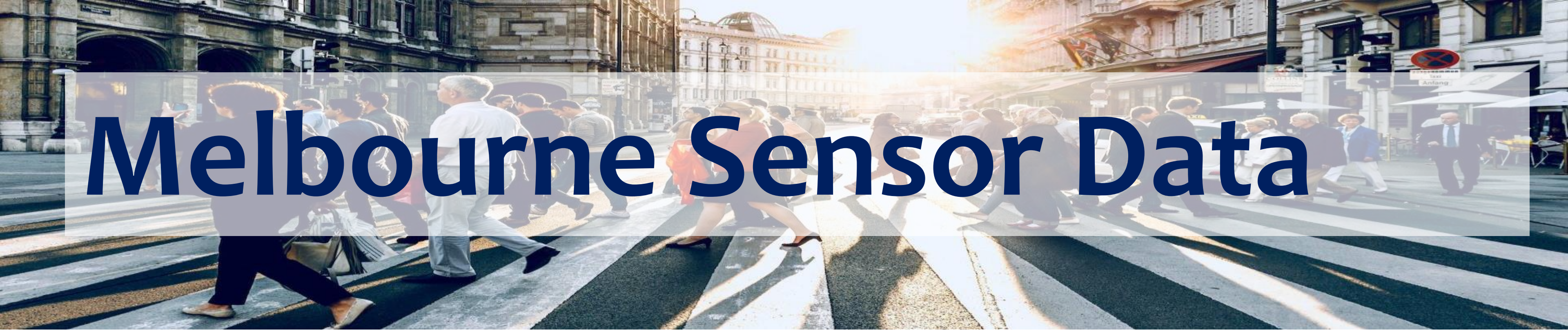


- Network of pedestrian sensors:
  - Record pedestrian counts every hour
  - Available openly at Melbourne Open Data portal

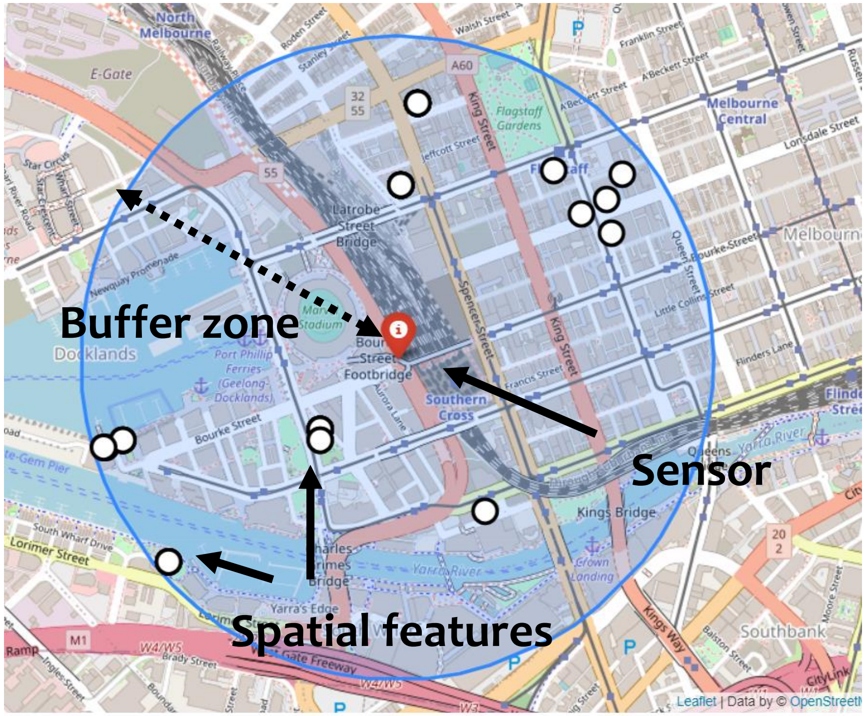
- Numerous additional open data sets, including:
  - Weather

1 https://www.data.vic.gov.au/discover-city-melbourne-open-data

UNIVERSITY OF LEEDS



Melbourne Sensor Data



Example buffer zone within which spatial features are linked to sensors

- Network of pedestrian sensors:
  - Record pedestrian counts every hour
  - Available openly at Melbourne Open Data portal
- Numerous additional open data sets, including:
  - Weather
  - Street furniture (benches, bins etc)
  - Buildings and landmarks

1 https://www.data.vic.gov.au/discover-city-melbourne-open-data

UNIVERSITY OF LEEDS

Melbourne Open Data

80 explanatory variables

Pedestrian count	Sensor ID	Hour of day	Day of week	Month of year	Number of nearby trees	Number of nearby offices	Number of nearby schools	Rainfall?
25	1	1	1	7	15	1	0	Yes
27	1	2	1	7	15	1	0	No
67	2	1	1	7	2	8	1	Yes
69	2	2	1	7	2	8	1	Yes

…

… 4 million sensor records

UNIVERSITY OF LEEDS

Melbourne Open Data

Training data...

Pedestrian count	Sensor ID	Hour of day	Day of week	Month of year	Number of nearby trees	Number of nearby offices	Number of nearby schools	Rainfall?
25	1	1	1	7	15	1	0	Yes
27	1	2	1	7	15	1	0	No
67	2	1	1	7	2	8	1	Yes
69	2	2	1	7	2	8	1	Yes

Testing data...

Pedestrian count	Sensor ID	Hour of day	Day of week	Month of year	Number of nearby trees	Number of nearby offices	Number of nearby schools	Rainfall?
?	1	2	1	7	15	1	0	No
?	2	1	1	7	2	8	1	Yes
?	2	2	1	7	2	8	1	Yes

Model selection

- Considered two machine learning models (compared against linear regression)
- Evaluated accuracy using 10-fold cross-validation

Model	MAE	RMSE
Random Forest regression		
XGBoost		
Linear regression		

UNIVERSITY OF LEEDS

Model selection

- Considered two machine learning models (compared against linear regression)
- Evaluated accuracy using 10-fold cross-validation

Error metrics

Model	MAE	RMSE
Random Forest regression	89.88	179.62
XGBoost	121.35	207.40
Linear regression	268.40	370.54

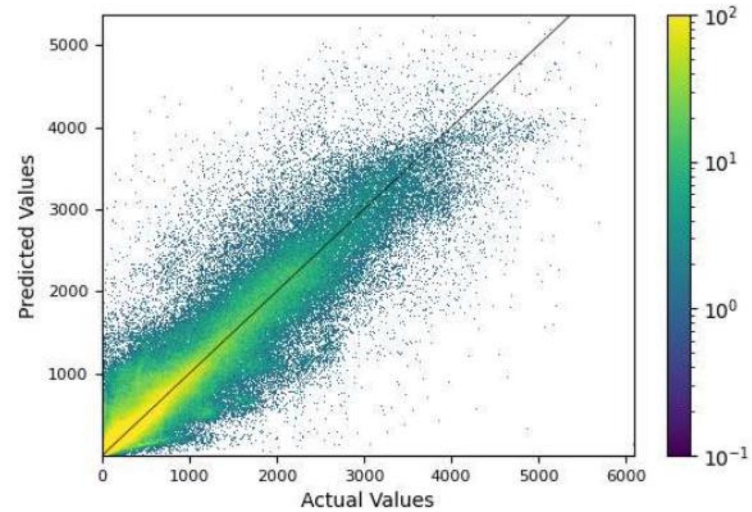
Random forest regressor selected as best performing model

UNIVERSITY OF LEEDS

Model evaluation

Random forest regressor selected as best performing model

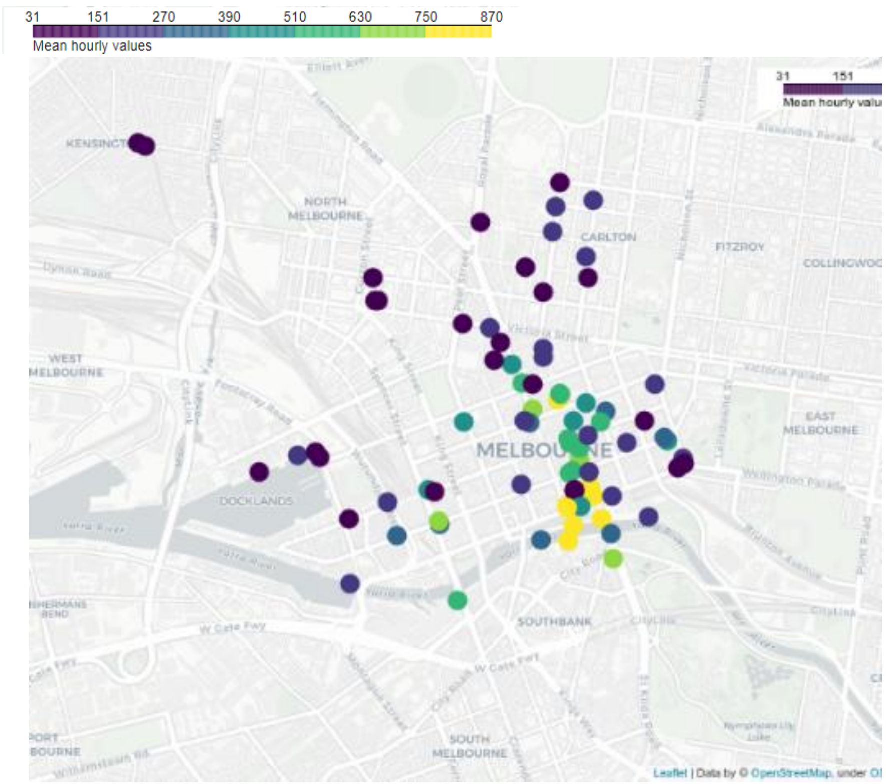
- Predicted counts-per-hour of pedestrians plotted against actual values from the sensor data
- Most predictions fall around diagonal (x=y), giving confidence model is not biased towards smaller or larger counts



MAE	RMSE
89.88	179.62

UNIVERSITY OF LEEDS

Model evaluation: spatial



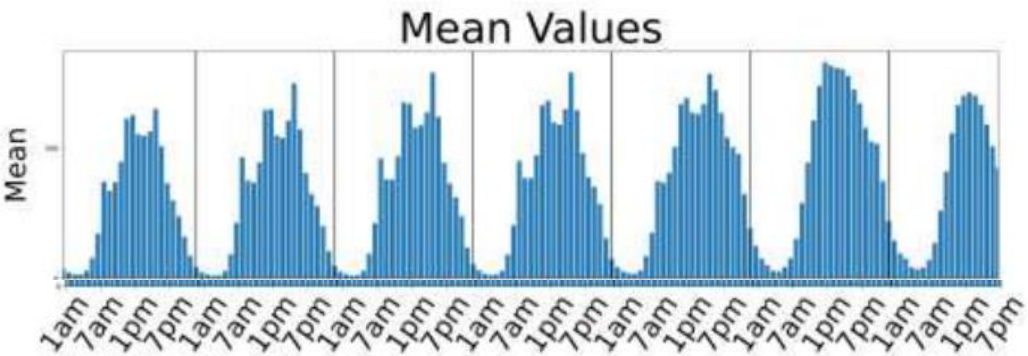
Mean

Central and southern sensors capture highest footfall

MAPE

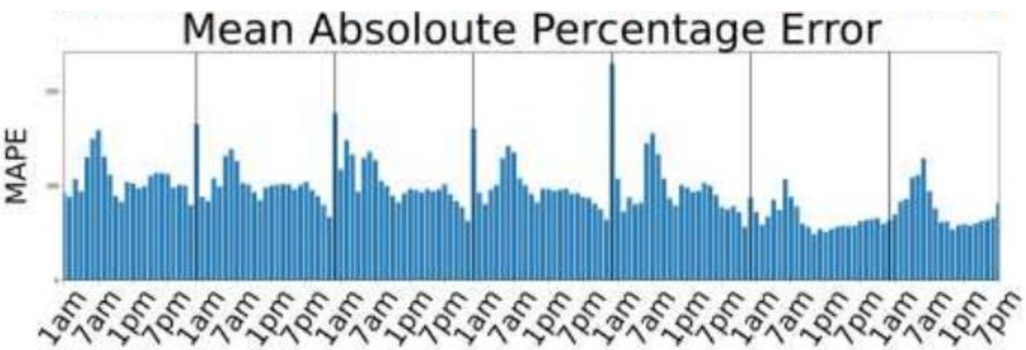
Several sensors with much larger percentage error

Model evaluation: temporal



Mean

Reflect typical city centre patterns

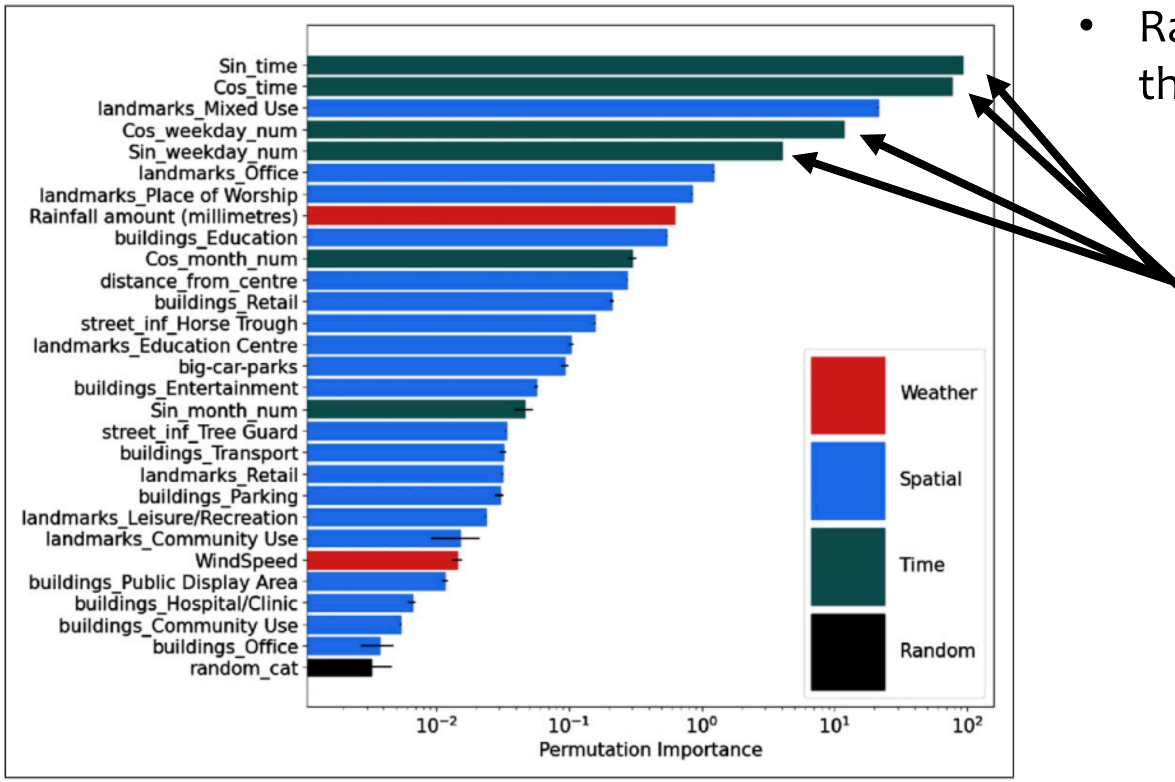


MAPE

Largest errors at night

UNIVERSITY OF LEEDS

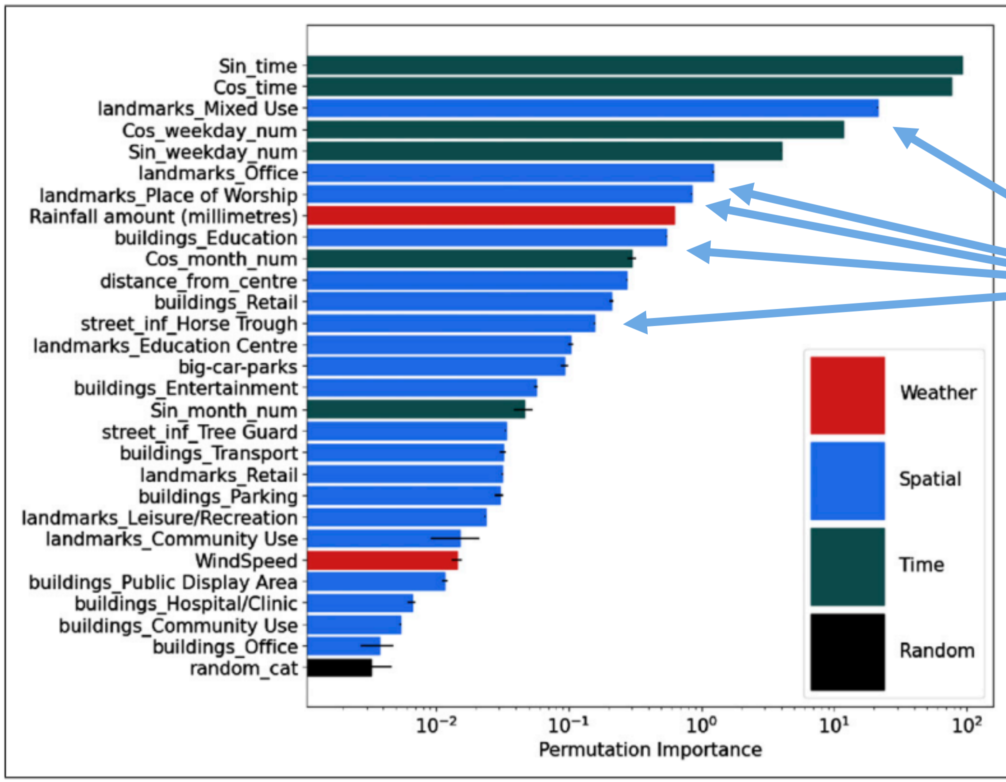
Feature importance



- Ranks features contribution to the model's predictions

UNIVERSITY OF LEEDS

Feature importance

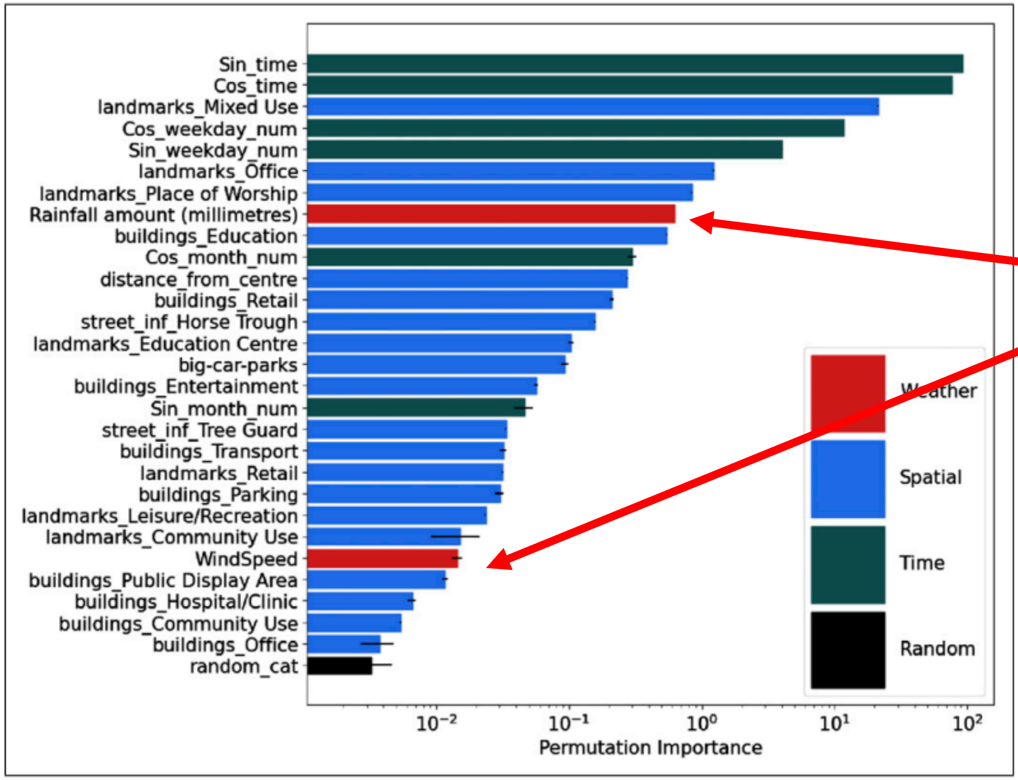


- Ranks features contribution to the model's predictions

Spatial variables

UNIVERSITY OF LEEDS

Feature importance

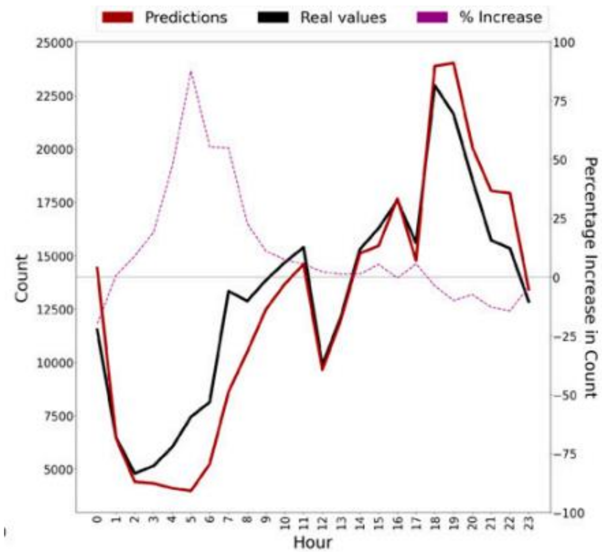


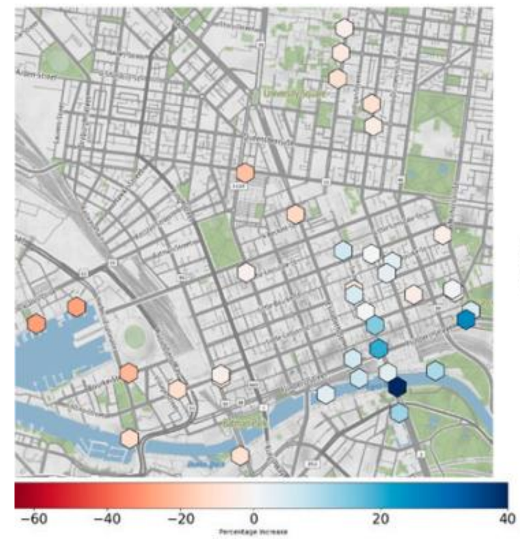
- Ranks features contribution to the model’s predictions

Weather variables

Evaluating events

- Model can be used as a tool to evaluate success of events
- E.g. Anzac Day Parade:
    - 5% more footfall in whole city over 24h
    - 72% more footfall from 3-10am
    - 128% more footfall at a sensor in south-east near parade location





UNIVERSITY OF LEEDS

Conclusions

- Our work:
  - accurately predict the number of pedestrians in time and space at un-sampled locations under different conditions
  - better understand the impact of the built environment and other contextual factors on pedestrian counts
  - Evaluate the success of past events

- Model performs reasonably well overall

- Some spatial and temporal variations in prediction error

- Beginning to make inferences about impact of urban environment

UNIVERSITY OF LEEDS

Thank you and questions

Published work: Asher, M., Oswald, Y. and Malleson, N., 2025. Understanding pedestrian dynamics using machine learning with real-time urban sensors. *Environment and Planning B: Urban Analytics and City Science*, p.23998083251319058.

UNIVERSITY OF LEEDS