Summary

The Urban Co-Creation Data Lab (UCD Lab) project aimed to support decision-making at the municipality level to provide citizens with high quality services in the areas of micromobility, waste management, parking, pollution, and emergency. The project aimed at developing a new generation of public services in the context of smart cities exploiting supercomputing facilities and public and private data to analyse complex combinations of large datasets in areas of public interest. The analytical model presented in this document was developed for the city of Lisbon regarding micromobility and was made publicly available to any interested person or institution. The UCD Lab was co-financed by CEF Telecom, the EU instrument to facilitate cross-border interaction between public administrations, businesses and citizens, and the project beneficiaries were: Universidade Nova de Lisboa, Município de Lisboa, Agência para a Modernização Administrativa, I.P., NEC Portugal - Telecomunicações e Sistemas, S.A, and Barcelona Supercomputing Center - Centro Nacional de Supercomputación.

Service description

This service allows to predict the bike docks occupation ratio for 84 bike stations of the docked bike sharing service in Lisbon (GIRA), for every 3 hours during a week.

Analytical model

Input data and variables

In Table 1 are presented the datasets necessary to develop the analytical model for #1 Micromobility use case.

Table 1. Datasets necessary for the development of the analytical model for #1 Micromobility use case.

Dataset	Source	Open data
Bike docks occupation (GIRA service)	Lisboa Aberta	Yes
Weather data	Instituto Português do Mar e da Atmosfera (IPMA)	No

The variables necessary to run the analytical model that serve as the basis of the service are presented in Table 2.

Table 2. Input data description of the micromobility model necessary to run the service.

Variable	Description	Type
datetime	Timestamp (hour)	DATE (dd/mm/yyyy HH:mm)
station	Unique id of the GIRA station	INTEGER
station_designation	Bike station designation	STRING
Id_weather	Weather station id	INTEGER
MEAN_numdocasvacias	Mean number of empty docks in periods of 3 hours	FLOAT

num_docks	Number of docks in the bike station	INTEGER
temp	Mean temperature (ºC) recorded in a period of 3 hours	FLOAT
precip	Mean accumulated precipitation (mm) in a period of 3 hours	FLOAT
Is_business_day	Binary flag identifying business days and weekends	INTEGER

Model

The analytical models that support the service is a Seasonal Auto-regressive Integrated Moving Average with Exogenous Factors (SARIMAX) (Box et al., 2008) model. For each bike station a SARIMAX model was trained with data from 1/09/2020 to 23/11/2020 and tested with data from 24/11/2020 to 30/11/2020, using the optimized parameters at bike station level. Temperature, precipitation and the indication of business days and weekends were used as exogenous variables.

Output

In Table 3 is presented the description of the output data provided by the analytical model.

Table 3. Output data description for the micromobility model.

Column	Description	Type
station	Unique id of the bike station	INTEGER
date	3 hours period for the prediction	DATE (yyyy-mm-dd HH:mm:ss)
predicted_station	Predicted number of empty docks in a period of 3 hours	FLOAT

The predicted occupation ratio was derived from the prediction of the number of empty docks. The predicted inoccupation ratio was computed dividing the predicted number of empty docks by the total number of docks of the bike station. The predicted occupation ratio was computed subtracting the predicted inoccupation ratio by one.

Evaluation

The model's quality for each bike station was assessed through the computation of the root mean square error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE) (de Myttenaere et al., 2016). In Figure 1 is presented the number of bike stations belonging to each forecast quality category derived from Lewis (1982), based on MAPE values. In Table 4 are presented the quality measures computed for each bike station.

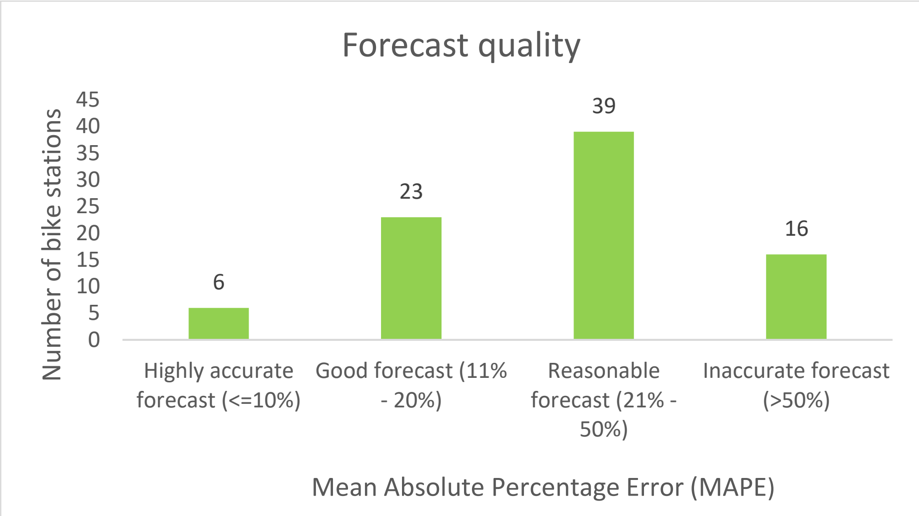

Figure 1. Number of bike stations in each class regarding forecast quality based on the scale developed by Lewis (1982).

Table 4. Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) obtained for each bike station in the testing sample of the SARIMAX model.

station	order	seasonal order	RMSE	MAE	MAPE (%)
101	(1, 0, 0)	(0, 0, 0, 56)	5,4	5,0	38
102	(3, 0, 0)	(0, 0, 0, 56)	4,3	3,6	59
103	(3, 0, 3)	(0, 0, 0, 56)	5,3	5,0	27
104	(1, 0, 1)	(0, 0, 0, 56)	3,6	3,3	9
105	(3, 0, 1)	(2, 0, 0, 56)	3,2	2,8	8
106	(1, 0, 1)	(0, 0, 2, 56)	2,6	2,4	18
107	(3, 0, 1)	(1, 0, 0, 56)	2,1	1,7	14
108	(1, 0, 1)	(0, 0, 0, 56)	2,0	1,8	14
109	(1, 0, 1)	(2, 0, 0, 56)	3,6	3,1	33
110	(1, 0, 1)	(2, 0, 0, 56)	3,7	3,0	22
112	(1, 0, 1)	(2, 0, 0, 56)	3,5	3,0	58
113	(2, 0, 0)	(2, 0, 0, 56)	3,8	3,4	48
114	(3, 0, 0)	(2, 0, 0, 56)	2,3	2,0	22
115	(3, 0, 1)	(1, 0, 2, 56)	2,1	1,8	24
208	(1, 0, 1)	(1, 0, 0, 56)	3,6	3,0	18
212	(2, 0, 2)	(0, 0, 0, 56)	3,8	2,9	27
214	(5, 0, 0)	(0, 0, 0, 56)	3,7	2,9	18
215	(1, 0, 1)	(0, 0, 1, 56)	2,6	2,2	16
216	(3, 0, 1)	(1, 0, 0, 56)	3,5	3,3	22
218	(3, 0, 0)	(2, 0, 0, 56)	1,9	1,5	10
219	(3, 0, 1)	(1, 0, 1, 56)	4,1	3,5	22
222	(4, 0, 3)	(1, 0, 0, 56)	7,2	6,1	64

224	(5, 0, 1)	(2, 0, 0, 56)	4,7	4,1	29
225	(5, 0, 1)	(1, 0, 0, 56)	4,1	3,7	45
301	(1, 0, 1)	(2, 0, 0, 56)	4,4	3,5	43
304	(1, 0, 1)	(1, 0, 2, 56)	1,9	1,6	14
305	(4, 0, 2)	(2, 0, 0, 56)	2,1	1,9	14
306	(3, 0, 1)	(2, 0, 0, 56)	1,8	1,6	11
307	(1, 0, 1)	(2, 0, 0, 56)	4,1	3,4	12
308	(1, 0, 1)	(1, 0, 0, 56)	2,3	1,8	40
309	(2, 0, 0)	(2, 0, 0, 56)	2,6	2,2	89
401	(2, 0, 0)	(2, 0, 0, 56)	1,5	1,1	9
403	(2, 0, 0)	(2, 0, 0, 56)	2,7	2,2	12
406	(1, 0, 1)	(0, 0, 0, 56)	2,3	1,8	14
407	(5, 0, 2)	(1, 0, 0, 56)	3,1	2,6	23
408	(1, 0, 1)	(0, 0, 0, 56)	2,1	1,8	28
410	(1, 0, 1)	(2, 0, 0, 56)	4,2	3,5	12
412	(1, 0, 1)	(1, 0, 0, 56)	1,9	1,6	11
413	(1, 0, 4)	(0, 0, 2, 56)	2,2	1,6	38
414	(3, 0, 1)	(0, 0, 0, 56)	4,0	3,5	32
415	(1, 0, 1)	(0, 0, 0, 56)	1,8	1,5	9
416	(1, 0, 2)	(0, 0, 1, 56)	0,9	0,8	5
417	(2, 0, 0)	(0, 0, 0, 56)	3,3	2,7	15
419	(2, 0, 0)	(2, 0, 0, 56)	5,5	4,6	41
420	(3, 0, 1)	(2, 0, 0, 56)	3,7	3,2	14
421	(2, 0, 0)	(2, 0, 0, 56)	6,0	5,1	68
423	(1, 0, 0)	(1, 0, 0, 56)	1,7	1,3	26
426	(3, 0, 4)	(0, 0, 0, 56)	2,3	2,0	21
427	(1, 0, 1)	(0, 0, 1, 56)	1,9	1,6	19
428	(1, 0, 1)	(2, 0, 0, 56)	1,9	1,5	16
430	(2, 0, 0)	(0, 0, 0, 56)	2,2	1,8	13
431	(3, 0, 1)	(0, 0, 2, 56)	3,3	2,7	55
432	(3, 0, 1)	(0, 0, 0, 56)	2,1	1,7	21
433	(2, 0, 0)	(2, 0, 0, 56)	2,9	2,2	57
442	(3, 0, 0)	(1, 0, 0, 56)	2,8	2,4	28
443	(1, 0, 1)	(0, 0, 0, 56)	3,1	2,2	36
446	(3, 0, 0)	(1, 0, 0, 56)	9,1	7,6	47
449	(4, 0, 3)	(0, 0, 0, 56)	3,1	2,6	48
450	(2, 0, 2)	(2, 0, 0, 56)	2,4	2,0	75
452	(2, 0, 0)	(2, 0, 0, 56)	2,4	2,0	55
453	(3, 0, 1)	(2, 0, 0, 56)	2,2	1,8	19
456	(3, 0, 0)	(0, 0, 0, 56)	7,0	6,1	33
457	(1, 0, 1)	(2, 0, 0, 56)	2,5	1,7	103
459	(1, 0, 0)	(0, 0, 0, 56)	5,7	4,4	127
460	(2, 0, 0)	(1, 0, 0, 56)	3,0	2,3	49
462	(3, 0, 1)	(2, 0, 0, 56)	3,9	3,4	22
463	(3, 0, 0)	(0, 0, 0, 56)	2,6	2,1	86

464	(3, 0, 1)	(0, 0, 0, 56)	3,1	2,7	136
468	(1, 0, 1)	(0, 0, 0, 56)	6,4	5,1	202
471	(1, 0, 1)	(0, 0, 0, 56)	3,4	3,3	97
472	(1, 0, 1)	(0, 0, 0, 56)	3,9	3,0	53
473	(2, 0, 0)	(0, 0, 1, 56)	1,5	1,2	18
474	(3, 0, 1)	(2, 0, 0, 56)	2,5	2,1	27
475	(1, 0, 1)	(2, 0, 0, 56)	4,6	3,3	11
476	(1, 0, 1)	(2, 0, 0, 56)	3,1	2,5	17
480	(1, 0, 1)	(0, 0, 2, 56)	4,9	3,9	46
481	(1, 0, 1)	(0, 0, 0, 56)	2,9	2,3	26
483	(2, 0, 0)	(1, 0, 0, 56)	2,6	2,2	30
484	(3, 0, 0)	(1, 0, 0, 56)	4,5	3,9	37
485	(1, 0, 1)	(1, 0, 1, 56)	2,6	2,2	45
486	(2, 0, 2)	(1, 0, 2, 56)	2,5	2,1	30
487	(2, 0, 3)	(1, 0, 0, 56)	1,9	1,6	24
488	(1, 0, 1)	(1, 0, 0, 56)	3,8	3,2	22
490	(2, 0, 0)	(1, 0, 0, 56)	3,5	2,6	32

Service

A dashboard was developed with the information provided by the predictions using the above-described model. The dashboard report on the predictions is presented in Figure 2. Through this report is possible to have a prediction about the bike docks occupation ratio for the next three hours for a specific bike station. The report as also a visual with the occupation ratio from the last 30 updates recorded by the system. The predicted occupation is also presented in a map where each bike station is coloured accordingly with the predicted value of occupation ratio. In the report was implemented a visual with warning signs that communicates to the operator if the occupation ratio for a bike station is low (in this case could be the need to rebalance the bike station with more bikes if the occupation ratio is below 30%) to guarantee that a user of the service has available bikes to pick up. On the contrary, the warning system also alerts the operator if a bike station is very occupied (it was considered an occupation ratio above 70%). In this case could be the need to remove some bikes of the bike station, to guarantee that a user of the service has available docks to drop off a bike.

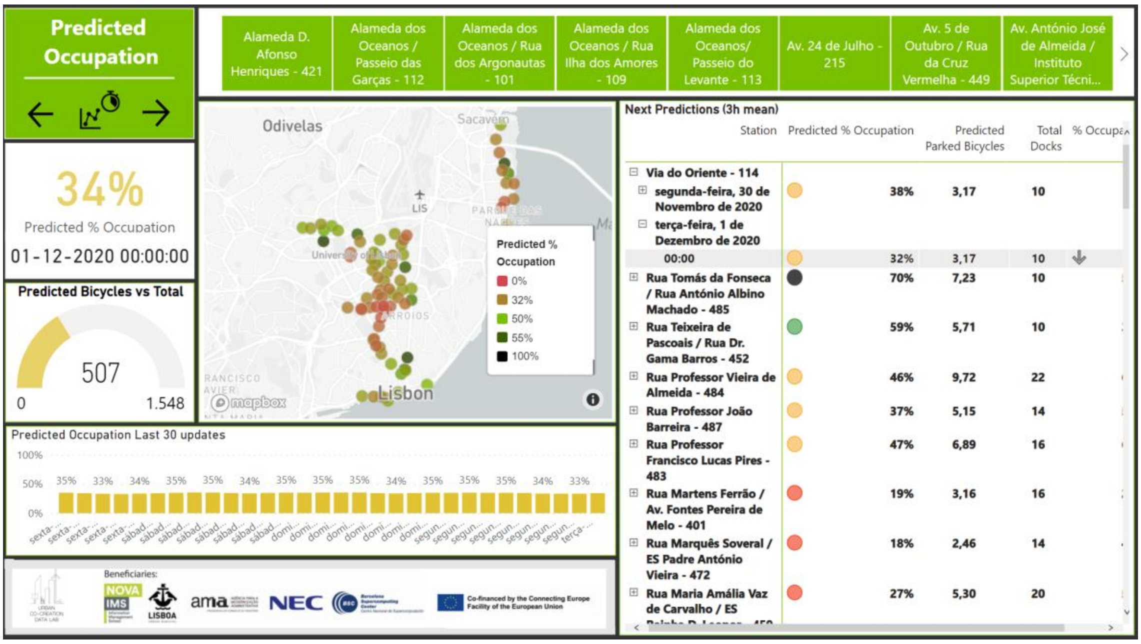

Figure 2. Report of predicted bike occupation ratio for #1 Micromobility use case.

References

Box, G., Jenkins, G., & Reinsel, G. (2008). *Time Series Analysis* (4th Editio). Hoboken.

de Myttenaere, A., Golden, B., Le Grand, B., & Rossi, F. (2016). Mean Absolute Percentage Error for regression models. *Neurocomputing*, *192*, 38–48. https://doi.org/10.1016/j.neucom.2015.12.114

Lewis, C. D. (1982). *Industrial and business forecasting methods*. Butterworths.