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 parking 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 identify the probability of illegal parking for a specific road segment and period of day.

Analytical model

Input data

In Table 1 are presented the datasets necessary to develop the analytical model for #3 Parking use case.

Table 1. Datasets necessary for the development of the analytical model for #3 Parking use case.

Dataset	Source	Open data
Parking illegalities	Lisbon City Council (CML) – Municipal Police	No
Weather data	Portuguese Institute for Sea and Atmosphere (IPMA)	No
Roads	Lisbon City Council (CML)	No
Waze jams	Lisbon City Council (CML)	No
Radars	Lisbon City Council (CML)	Yes
Traffic light areas	Lisbon City Council (CML)	Yes
Public hospitals	Lisbon City Council (CML)	Yes
Private hospitals	Lisbon City Council (CML)	Yes
Health centers	Lisbon City Council (CML)	Yes
Public schools - Pre-schools	Lisbon City Council (CML)	Yes
Public schools – 1st cycle schools	Lisbon City Council (CML)	Yes

Public schools – 2nd cycle schools	Lisbon City Council (CML)	Yes
Public schools – 3rd cycle schools	Lisbon City Council (CML)	Yes
Public schools – secondary schools	Lisbon City Council (CML)	Yes
Private schools -Pre-schools	Lisbon City Council (CML)	Yes
Private schools – 1st cycle schools	Lisbon City Council (CML)	Yes
Private schools – 2nd and 3rd cycle schools	Lisbon City Council (CML)	Yes
Private schools – secondary schools	Lisbon City Council (CML)	Yes
Faculties, schools and institutes	Lisbon City Council (CML)	Yes
Train stations	Lisbon City Council (CML)	Yes
Metro stations	Lisbon City Council (CML)	Yes
Bus stations	CARRIS	No

Modelling

For the development of the analytical model for #3 Parking, data of illegal parking occurrences provided by Lisbon Municipal Police from 02/01/2017 to 31/12/2020 was used. First, a text classification model was implemented to classify the description of each parking illegality into one of four classes – on crosswalk, on sidewalk, conditions access, reserved for the disabled, reserved places, others (when the description it does not fit any of the other classes) and unknown (when there is no description). The text classification model was based on a multi-class logistic regression that receives a DistilBERT (Sanh et al., 2019) vector representation of each description and gives a probability of that description belonging to one of the classes, being the class with the highest probability the one chosen. After retrieving the illegality class, the modelling strategy developed for the parking use case was divided in two stages. In the first stage the probability of the occurrence of illegal parking was computed. In Table 2 are presented the variables necessary for the development of the first stage of the modelling strategy.

Table 2. Variables used for the computation of parking illegalities probability.

Variable	**Description**	**Type**
road_id	Unique identifier of the road segment	INTEGER
temperature	Code for temperature recorded during a specific day period: 10=]∞, 10 °C]; 20=]10 °C – 20 °C]; 30=]20 °C – 30 °C]; 40=]30 °C – 40 °C]	INTEGER

precipitation	Code for precipitation recorded during a specific day period: 0.01=[0 mm – 0,01 mm]; 2.5=]0,01 mm – 2,5 mm]; 5=]2,5 mm – 5 mm]; 10=]5 mm - ∞[	INTEGER
period	Code to identifying the period of day: 1=[0h – 4h[; 2=[4h – 7h[; 3=[7h – 10h[; 4=[10h – 14h[; 5=[14h – 17h[; 6=[17h – 20h[; 7=[20h – 24h[	INTEGER
off_day	Flag identifying weekends and holidays: 0=business day; 1=weekend or holiday	INTEGER
class_parking	Parking illegalities classification: ‘condiciona_acessos’, ‘deficientes’, ‘lugares_reservados’, ‘no_passeio’, ‘outras’, ‘passadeira’, ‘desconhecido’.	STRING
count	Group by count of the combination of [road_id], [temperature], [precipitation], [period], [off_day] and [illegalities_classification]	INTEGER
sum_illegalities	Sum of parking illegalities for the possible combinations of [road_id], [temperature], [precipitation], [period], [off_day], and [parking illegalities classification]	INTEGER

The probability of parking illegalities was computed dividing [sum_illegalities] by [count].

In the second stage of the modelling strategy, all combinations of the features [road_id], [temperature], [precipitation], [period], [off_day], and [class_parking] with a count value lower than 100 were discarded, as they were not considered statistically significant. To estimate a probability for the cases where statistical significance was not met a machine learning algorithm, namely LightGBM (LGBM) (Lv et al., 2021) was used. LGBM is a gradient boosting framework that uses tree-based learning algorithms. This framework was implemented in two different steps: 1) in which was used as a classification algorithm to identify (for the situations in which the combination of features was < 100) the observations were the probability was non null; and 2) from the identified observations in the previous step, LGBM was used as a regressor to assign a probability of the occurrence of illegal parking for each observation.

In Table 3 is presented the input data that was used for the application of the classification and regression algorithm.

Table 3. Input data for the prediction of the probability of traffic accidents occurrences in the observations, were the combination of the features [road_id], [temperature], [precipitation], [period], [off_day] and [class_parking] is < 100.

Variable	**Description**	**Type**
road_id	Unique identifier of the road segment	INTEGER
road_name	Road name	STRING
is_off_day	Flag identifying weekends and holidays: 0=business day; 1=weekend or holiday	INTEGER
temperature	Code for temperature recorded during a specific day period: 10=]∞, 10 °C]; 20=]10 °C – 20 °C]; 30=]20 °C – 30 °C]; 40=]30 °C – 40 °C]	INTEGER

precipitation	Code for precipitation recorded during a specific day period: 0.01=[0 mm – 0,01 mm]; 2.5=]0,01 mm – 2,5 mm]; 5=]2,5 mm – 5 mm]; 10=]5 mm - ∞[	INTEGER
condiciona acessos	Flag identifying the parking illegality category “condition accesses”	INTEGER
lugar de deficientes	Flag identifying the parking illegality category “handicapped parking space”	INTEGER
lugares reservados	Flag identifying the parking illegality category “reserved parking space”	INTEGER
na passadeira	Flag identifying the parking illegality category “crosswalk”	INTEGER
no passeio	Flag identifying the parking illegality category “sidewalk”	INTEGER
outras	Flag identifying the parking illegality category “others”	INTEGER
day_period_1	Flag identifying the period of day [0h – 4h[	INTEGER
day_period_2	Flag identifying the period of day [4h – 7h[	INTEGER
day_period_3	Flag identifying the period of day [7h – 10h[	INTEGER
day_period_4	Flag identifying the period of day [10h – 14h[	INTEGER
day_period_5	Flag identifying the period of day [14h – 17h[	INTEGER
day_period_6	Flag identifying the period of day [17h – 20h[	INTEGER
day_period_7	Flag identifying the period of day [20h – 24h[	INTEGER
waze_proxy	Sum of jams in a road segment in all historical period	INTEGER
lane_number	Number of road lanes	INTEGER
vel_max	Maximum velocity allowed in a road segment	INTEGER
comp	Road segment length	FLOAT
semaforo	Flag identifying if the road segment is in a traffic light area: 1=road segment in a traffic light area	INTEGER
COUNT_hospitals	Count of hospitals in the nearest road segment	INTEGER
COUNT_health_centers	Count of health centers in the nearest road segment	INTEGER
COUNT_schools0	Count of pre-schools in the nearest road segment	INTEGER
COUNT_schools1	Count of 1st cycle schools in the nearest road segment	INTEGER
COUNT_schools_2_3	Count of 2nd and 3rd cycle schools in the nearest road segment	INTEGER
COUNT_schools_12	Count of secondary schools in the nearest road segment	INTEGER
COUNT_universities	Count of universities in the nearest road segment	INTEGER
COUNT_train	Count of train stations in the nearest road segment	INTEGER
COUNT_metro	Count of metro stations in the nearest road segment	INTEGER

COUNT_bus	Count of bus stations in the nearest road segment	INTEGER

The LGBM model used for classification, in terms of overall quality, has an Area Under Curve (AUC) (Huang & Ling, 2005) of 0,88. The LGBM model used for regression was assessed through the computation of the Mean Absolute Percentage Error (MAPE) (de Myttenaere et al., 2016) having an error of 53%.

Output data

The output data of the models corresponds to the probability of illegal parking grouped by the variables considered [road_name], [is_off_day], [temperature], [precipitation], [day_period] (Table 4).

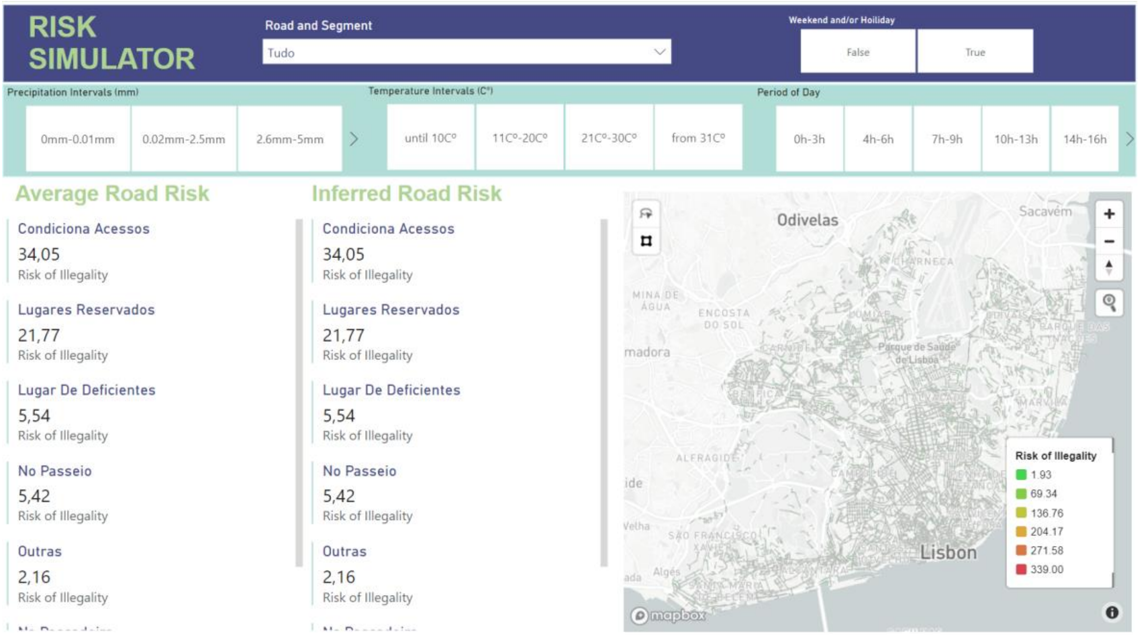
Table 4. Output data of the analytical model developed in the use case #3 Parking.

Variable	Description	Type
road_name	Unique identifier of the road segment	INTEGER
is_off_day	Flag identifying weekends and holidays: 0=business day; 1=weekend or holiday	INTEGER
temperature	Temperature intervals (ºC): [<10]; [11 – 20]; [21 – 30]	INTEGER
precipitation	Precipitation intervals (mm): [0 – 0.01]; [0.02 – 2.5]; [2.6 – 5]	INTEGER
day_period	Time intervals (h): [0 – 3]; [4 – 6]; [7 – 9]; [10 – 13]; [14 – 16]; [17 – 19]; [20 – 23].	INTEGER
risk_illegal_parking	Probability of the occurrence of illegal parking in a road segment by illegal parking category	FLOAT

To allow a better comprehension of the probability of illegal parking, the probability was multiplied by 100 000.

Service

A report (Figure 1) was implemented that allows to predict the risk of illegal parking by category, considering the following dimensions: 1) weekend and/or holiday; 2) precipitation; 3) temperature; and 4) period of day. The risk of illegal parking is possible to identify for all the city or for a specific road segment considering the above-mentioned dimensions.



RISK SIMULATOR

Road and Segment: Tudo

Weekend and/or Hoiliday: False | True

Precipitation Intervals (mm): 0mm-0.01mm | 0.02mm-2.5mm | 2.6mm-5mm

Temperature Intervals (C°): until 10C° | 11C°-20C° | 21C°-30C° | from 31C°

Period of Day: 0h-3h | 4h-6h | 7h-9h | 10h-13h | 14h-16h

Average Road Risk

Condiciona Acessos
34,05
Risk of Illegality

Lugares Reservados
21,77
Risk of Illegality

Lugar De Deficientes
5,54
Risk of Illegality

No Passeio
5,42
Risk of Illegality

Outras
2,16
Risk of Illegality

Inferred Road Risk

Condiciona Acessos
34,05
Risk of Illegality

Lugares Reservados
21,77
Risk of Illegality

Lugar De Deficientes
5,54
Risk of Illegality

No Passeio
5,42
Risk of Illegality

Outras
2,16
Risk of Illegality

Figure 1. Illegal parking risk simulator report.

References

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

Huang, J., & Ling, C. X. (2005). Using AUC and Accuracy in Evaluating Learning Algorithms. *IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING*, *17*(3), 299–310.

Lv, Z., Lou, R., Feng, H., Chen, D., & Lv, H. (2021). Novel Machine Learning for Big Data Analytics in Intelligent Support Information Management Systems. *ACM Trans. Manage. Inf. Syst.*, *13*(1). https://doi.org/10.1145/3469890

Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). *DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter*. 2–6. http://arxiv.org/abs/1910.01108