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 emergency 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 risk of a traffic accident that involved the intervention of firefighters 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 #5 Emergency use case.

Table 1. Datasets necessary for the development of the analytical model for #5 Emergency use case.

Dataset	Source	Open data
Traffic accidents occurrences	Lisbon City Council (CML) – Lisbon firefighters	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

Modelling

For the development of the analytical model for the #5 Emergency use case, data of traffic accidents that required firefighters' intervention from 01/01/2013 to 15/12/2020 was used. The modelling strategy developed for the emergency use case was divided in two stages. In the first stage the probability of occurrence of a traffic accident by road segment for a specific day period, considering meteorological conditions (namely temperature and precipitation) was computed. In Table 2 are presented, the variables required for the development of the first stage of the modelling strategy.

Table 2. Variables used for the computation of traffic accidents 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
count	group by count of the combination of [road_id], [temperature], [precipitation], [period], and [off_day]	INTEGER
sum_occurrences	Sum of traffic accidents occurrences for the possible combinations of [road_id], [temperature], [precipitation], [period], and [off_day]	INTEGER

The probability of traffic accidents occurrence was computed dividing [sum_occurrences] by [count].

In the second stage of the modelling strategy, all combinations of the features [road_id], [temperature], [precipitation], [period], and [off_day], with a 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 traffic accident for each observation.

In Table 3 one can see the variables that were 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], and [off_day] 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
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
radar	Number of radars in a road segment	INTEGER
semaforo	Flag identifying if the road segment is in a traffic light area: 1=road segment in a traffic light area	INTEGER
probability	Probability of the occurrence of a traffic accident in a road segment	FLOAT

The LGBM model used for classification, in terms of overall quality, has an Area Under Curve (AUC) (Huang & Ling, 2005) of 0,80. 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 20%.

Output data

The output data of the models corresponds to the probability/risk of traffic accident grouped by the variables [road_id], [lane_number], [vel_max], [radar], and [semaforo] (Table 4).

Table 4. Output data of the analytical model developed in the use case #5 – Emergency.

Variable	Description	Type
road_id	Unique identifier of the road segment	INTEGER
lane_number	Number of road lanes	INTEGER
vel_max	Maximum velocity allowed in a road segment	INTEGER
radar	Number of radars in a road segment	INTEGER
semaforo	Flag identifying if the road segment is in a traffic light area: 1=road segment in a traffic light area	INTEGER
risk_traffic_accident	Probability of existence of traffic accident (*100 000)	FLOAT

To allow a better comprehension of the risk of traffic accident, the probability was multiplied by 100 000.

Service

A report was developed with a traffic accidents simulator that evaluates the risk of traffic accidents by road segments (Figure 1). The operator can for a specific road segment, change the number of lanes, the top speed allowed, add or remove radar, and add or remove traffic lights. This will generate a predicted value of traffic accidents risk for a specific road segment. In this way emergency managers can simulate the impacts that changes in road infrastructure can have in the risk of traffic accidents that required firefighters intervention.

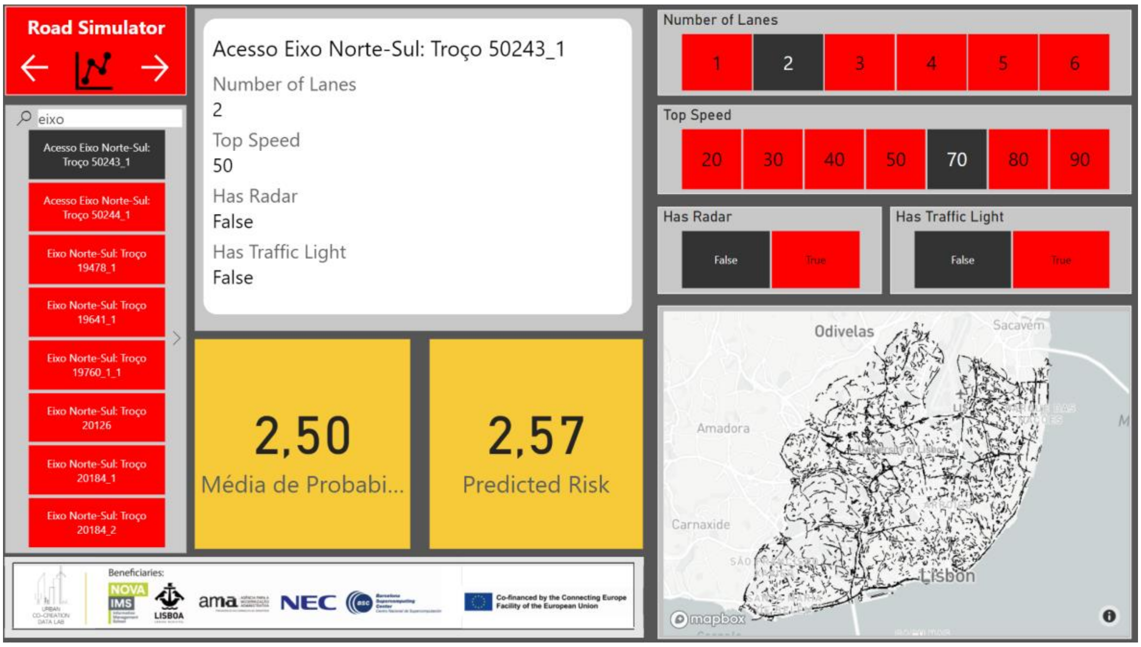

Figure 1. Traffic accidents risk simulator report.

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