DM
DEPARTAMENTO
DE MATEMÁTICA
TÉCNICO LISBOA

STATISTICAL METHODS IN DATA MINING

MEGIE, MECD, MMAC, OTHERS, 1st SEMESTER, 2023/2024

Project 1

Handed out on October 2, 2023.

To be **handed back** by **October 20**, 2023.

This study investigates a person's propensity to accept the coupon that is suggested to him in various driving situations. The objective was to comprehend how consumers reacted to suggestions from an in-car recommender system that offers discounts for local businesses.

The data was collected via a survey on Amazon Mechanical Turk and the survey describes different driving scenarios including the destination, current time, weather, passenger, etc., and then ask the person whether he will accept the coupon if he is the driver. For more information about the dataset, please refer to the paper: Wang, Tong, Cynthia Rudin, Finale Doshi-Velez, Yimin Liu, Erica Klampfl, and Perry MacNeille. 'A Bayesian framework for learning rule sets for interpretable classification.' The Journal of Machine Learning Research 18, no. 1 (2017): 2357-2393. The paper can be download here.

The researcher has rudimentary knowledge about Data Mining and wants your help to work on his/her problem.

You can grab the dataset: here.

Problem to solve

The objective of the prediction problem is to predict whether a client will accept a coupon for a specific venue, taking into account demographic and contextual variables.

Steps

1. Perform a exploratory data analysis and discuss what you have learned from this analysis.

2. Solve your classification problem. Consider several classification methods and discuss how can they contribute for the solution of your problem. Include in your discussion topics such as the options that you have made in building each classifier; interpretation of results; validation of the methods used and possible assumptions; advantages and disadvantages of each alternative; etc. Have in mind that some of the explanatory variables may be irrelevant

to the classification problem and that you may need to do some preprocessing methodologies of your data set e.g. dimensionality reduction techniques.

3. Imagine you are going to meet the researcher who contacted you. Report to him/her what you have learned about the problem. Discuss limitations of the analysis you have done and provide suggestions for future work.

About the groups:

- Students should organize themselves in groups of up to 5 persons.

About the report:

- You have to deliver a report including an explanation of what you have done, the visualization of our results, and a critical discussion of your results.

- The report should not exceed 12 pages.

- The **commented R/Python code** and the **report** must be uploaded to the Fénix webpage.