Université de Lorraine - UFR Math-Info
Master SCA – M1
2015-2016

# Fiche de projet tutoré / Project form

## Data-science challenge

**Encadrement / Supervisors**
Couceiro Miguel
Raïssi Chedy
Galbrun Esther

**Description / Description**
The aim of this project is for the students to take part in a data-science project. More specifically, the aim is to prepare and submit an entry to one of the competitions active on kaggle.com at the time, in order to appear on the leaderboard. The students will need to follow the submission and formatting rules associated with the competition.

**Informations diverses : matériel nécessaire, contexte de réalisation /**
**Various information: material, context of realization**
Students will be able to select their dataset(s) from kaggle's website at https://www.kaggle.com/. As an example, we cite the challenge :

***Predict the category of crimes that occurred in the city by the bay.***
*From 1934 to 1963, San Francisco was infamous for housing some of the world's most notorious criminals on the inescapable island of Alcatraz. Today, the city is known more for its tech scene than its criminal past. But, with rising wealth inequality, housing shortages, and a proliferation of expensive digital toys riding BART to work, there is no scarcity of crime in the city by the bay. From Sunset to SOMA, and Marina to Excelsior, this competition's dataset provides nearly 12 years of crime reports from across all of San Francisco's neighborhoods. Given time and location, you must predict the category of crime that occurred. We're also encouraging you to explore the dataset visually.*

**Livrables et échéancier / Deliverable and schedule**
The schedule depends on the timeline of the competition. One working submission should be made to the website before the end of the project (of course, students are free to revise and improve their solution until the end of the competition beyond the duration of the project, if they so wish).

**Bibliographie / References**

- Data Mining and Analysis: Fundamental Concepts and Algorithms, by M. J. Zaki and W. Meira Jr
- Pattern Recognition and Machine Learning by C. Bishop
- http://scikit-learn.org