My projects

Academic and personal projects

Gabin Drouard
gabin.drouard@gmail.com

More information about my CV and my professional activities:
https://www.linkedin.com/in/gabin-drouard-8b1310178

Forewords

You will find in this document the projects carried out in an academic or personal framework. The vast majority of these projects were carried out during my studies at the french National School for Statistics and Information Analysis (ENSAI, France) and the University of Rennes 1 (France). The evaluation in the framework of Rennes 1 being made exclusively from projects, it explains in part the great number of projects realized. The work carried out during internships or job positions is therefore not present (see CV). These projects are ordered chronologically and are more or less detailed. Basic information will be filled in (theme of the project, workload, year *etc.*). Note finally that the mention ***Available*** indicates the possibility of obtaining the project in a complete or partial way (contact by e-mail / LinkedIn). Abbreviations are presented in Table 1 to identify projects by theme.

Table 1: Abbreviations and references to study themes

Theme of study	Abbreviation	Some details	$n^{(*)}$
Mathematics	$\mathcal{M}$	mainly theoretical math	4
Statistics	$\mathcal{S}$	statistics, biostatistics, epidemiology, clinical studies, survival analysis & affiliates	17
Bioinformatics	$\mathcal{B}$	omics analysis / genetics	4
Machine & Deep Learning	$\mathcal{L}$	machine learning & deep learning	5
Computer science	$\mathcal{C}$	full programming projects (R/Python)	2
Physics	$\mathcal{P}$	mechanics, thermodynamics	1
Economics	$\mathcal{E}$	macro/microeconomics & social	3
Cartography	$\mathcal{G}$	basic R skills	1

$(*)$ *: n is the number of projects for which the theme appears.*

Contents

1 [$\mathcal{S}$,$\mathcal{M}$] **The integration of a vaccination campaign and prevention measures in deterministic and stochastic SIR models** **3**

2 [$\mathcal{S}$,$\mathcal{E}$] **Identification of adolescents profiles who are actors or victims of delinquency and their socio-cultural characteristics** **3**

3 [$\mathcal{L}$,$\mathcal{S}$] **Diagnosing heterogeneity within meta-analyses using machine learning methods** **3**

4 [$\mathcal{S}$] **Tisagenlecleucel in adult patients: a complementary study using Simon's Two-Stage design** **4**

5 [$\mathcal{S}$,$\mathcal{M}$] **Poisson modeling of the number of *del Norte* salamanders by Bayesian inference** **4**

6 [$\mathcal{S}$] **Longitudinal weight modeling of kidney transplant patients in chronic remission** **4**

7 [$\mathcal{S}$,$\mathcal{B}$] **Phenotypic diversity and aggressiveness of Human Hepatocellular Carcinomas** **5**

8 [$\mathcal{L}$,$\mathcal{B}$] **Methodology : a deep learning model to predict RNA-Seq expression of tumours from whole slide images** **5**

9 [$\mathcal{B}$,$\mathcal{S}$,$\mathcal{L}$] **Dynamic Programming as an alternative to Gaussian mixtures in DNA copy number segmentation** **5**

10 [$\mathcal{S}$,$\mathcal{L}$] **Early diagnosis of nosocomial infection in newborns** **5**

11 [$\mathcal{S}$] **PK-PD modeling of antibody concentration evolution in patients with advanced urothelial carcinoma** **6**

12 [$\mathcal{S}$] **Effect of Methotrexate on disease free survival in patients treated for hematologic cancer with bone marrow transplants** **6**

13 [$\mathcal{S}$] **Occupational exposure to pesticides and risk of endometriosis** **6**

14 [$\mathcal{B}$,$\mathcal{S}$,$\mathcal{L}$] **Prediction of censorship and events from omics data (SNP) for patients treated with chemotherapy** **6**

15 [$\mathcal{S}$,$\mathcal{M}$] **Forecasting death headcounts in metropolitan France using Bootstrap methods** **7**

16 [$\mathcal{S}$] **Study of the French effective reproduction number R_t and its evolution during the 1^{st} quarantine period.** **7**

17 [$\mathcal{M}$] **M/M/∞ service model applied to population dynamics as an alternative to conventional prey-predation models** **7**

18 [$\mathcal{S}$] **Study of antibiotic prescription times for three common diseases** **7**

19 [$\mathcal{C}$] **Creation of a Health application** **8**

20 [$\mathcal{G}$] **Cartographic representation of the French legislative elections of 2018** **8**

21 [$\mathcal{C}$] **Inventory management using Python** **8**

22 [$\mathcal{E}$] **The impact of advertising on consumption and on macro-economic variables** **8**

23 [$\mathcal{S}$,$\mathcal{E}$] **Science level of 15-year-old French students: the blatant manifestation of social reproduction through extracurricular practices** **9**

24 [$\mathcal{P}$] **Modeling of heating in the frame of bicycle brakes** **9**

---

1 [$\mathcal{S}$,$\mathcal{M}$] The integration of a vaccination campaign and prevention measures in deterministic and stochastic SIR models

Project size :	small/medium	**Year :**	2021
Availability :	yes	**Group project :**	no
Theme :	$\mathcal{S}$,$\mathcal{M}$	**Framework :**	academic work

Keywords: SIR modeling, Compartmental modeling, Epidemiology, R
Main results:
The objective of this project was to introduce SIR compartmental models with improvements such as the integration of a vaccination or prevention campaign within the equations. It was shown that a delay in a prevention campaign had a significant impact on the incidence rate. The purpose of this work is to illustrate the situation of covid-19 and to provide insights on decision making at the political level.

2 [$\mathcal{S}$,$\mathcal{E}$] Identification of adolescents profiles who are actors or victims of delinquency and their socio-cultural characteristics

Project size :	medium	**Year :**	2021
Availability :	yes	**Group project :**	no
Theme :	$\mathcal{S}$,$\mathcal{E}$	**Framework :**	academic work

Keywords: Dimension reduction (MFA), Clustering, Socio-cultural characteristics, R
Main results:
Multi-variate study aimed at establishing profiles of adolescents according to their bullying status. Highlighting of socio-cultural characteristics (racial, economic, social) associated with a stalker profile.

3 [$\mathcal{L}$,$\mathcal{S}$] Diagnosing heterogeneity within meta-analyses using machine learning methods

Project size :	small	**Year :**	2021
Availability :	yes	**Group project :**	no
Theme :	$\mathcal{L}$,$\mathcal{S}$	**Framework :**	academic work

Keywords: Cochran's $\mathcal{Q}$, DBSCAN, GMM, K-means, R
Main results:
The objective of this short project was to take a scientific article and give it a critical look. The source of heterogeneity is then investigated using machine learning methods (DBSCAN, GMM, K-means) to question the use of certain studies within the meta-analysis.

4 [$\mathcal{S}$] Tisagenlecleucel in adult patients: a complementary study using Simon's Two-Stage design

Project size :	medium	**Year :**	2021
Availability :	yes	**Group project :**	yes
Theme :	$\mathcal{S}$	**Framework :**	academic work

Keywords: Simon's Two-Stage design, Sensitivity analysis, R
Main results:
Reworking a scientific article to propose a new study or improvements to the intermediate analyses. We chose to integrate a Two-Stage design called Simon's Two-Stage design to optimize the number of patients in a new study. The final objective was to propose an improved study that would be less costly in terms of patients and would allow us to answer hypotheses drawn from the literature on Cart T-cells.

5 [$\mathcal{S}$,$\mathcal{M}$] Poisson modeling of the number of *del Norte* salamanders by Bayesian inference

Project size :	medium	**Year :**	2021
Availability :	yes	**Group project :**	no
Theme :	$\mathcal{S}$,$\mathcal{M}$	**Framework :**	academic work

Keywords: Bayesian inference, Poisson modeling, R
Main results:
Modeling the number of del Norte salamanders by a Poisson distribution with Bayesian inference. Comparison of frequentist models with (over-dispersive) Bayesian models. MCMC convergence and mixing diagnostics *etc.*

6 [$\mathcal{S}$] Longitudinal weight modeling of kidney transplant patients in chronic remission

Project size :	medium	**Year :**	2021
Availability :	yes	**Group project :**	no
Theme :	$\mathcal{S}$	**Framework :**	academic work

Keywords: Mixed-effect modeling, Missing data imputation, R
Main results:
Modeling changes in patient weight from imputed data. Use of mixed effects models and investigation of interactions between co-variates. It has been shown for example that age and gender are significantly associated with variations in patient weight after transplantation. Layout of the project as a scientific article.

7 [$\mathcal{S}$,$\mathcal{B}$] Phenotypic diversity and aggressiveness of Human Hepatocellular Carcinomas

Project size :	substantial work	**Year :**	2020-2021
Availability :	confidential	**Group project :**	yes
Theme :	$\mathcal{S}$,$\mathcal{B}$	**Framework :**	academic work

Keywords: Omics, Survival analysis, Cox, Score, R
Main results:
The objective of this project is to determine the genetic traits that will allow a better qualification of the 2-year survival of the operated patients. Use of Methylation, RNA-seq and clinical datasets. A better characterization of periportal hepatocellular carcinomas was expected and has been achieved.

8 [$\mathcal{L}$,$\mathcal{B}$] Methodology : a deep learning model to predict RNA-Seq expression of tumours from whole slide images

Project size :	medium	**Year :**	2020-2021
Availability :	yes	**Group project :**	yes
Theme :	$\mathcal{L}$,$\mathcal{B}$	**Framework :**	academic work

Keywords: Convolutional Neural Networks, RNA-seq
Main results:
The main objective is to study an article (*https://doi.org/10.1038/s41467-020-17678-4*) in depth in order to propose a methodological rewriting or contribution. The main goal is therefore to use the available resources in order to write a scientific article with a rigorous methodological structure.

9 [$\mathcal{B}$,$\mathcal{S}$,$\mathcal{L}$] Dynamic Programming as an alternative to Gaussian mixtures in DNA copy number segmentation

Project size :	small/medium	**Year :**	2021
Availability :	yes	**Group project :**	no
Theme :	$\mathcal{B}$,$\mathcal{S}$,$\mathcal{L}$	**Framework :**	academic work

Keywords: Dynamic Programming, Gaussian Mixture, Copy number, R
Main results:
Using Dynamic Programming to counter the problems of Gaussian mixtures in DNA segmentation. Use of simulated data to illustrate the limits of GMM and application of Dynamic Programming to real data (lung cancer).

10 [$\mathcal{S}$,$\mathcal{L}$] Early diagnosis of nosocomial infection in newborns

Project size :	small/medium	**Year :**	2020
Availability :	yes	**Group project :**	no
Theme :	$\mathcal{S}$,$\mathcal{L}$	**Framework :**	academic work

Keywords: Clinical epidemiology, Forecasting, SVM, R
Main results:
Comparison of biomarkers and their efficacy in predicting nosocomial diseases in neonates. Use of SVM type models for prediction as well as Monte-Carlo type methods to estimate the area under a curve.

11 [$\mathcal{S}$] PK-PD modeling of antibody concentration evolution in patients with advanced urothelial carcinoma

Project size :	small	**Year :**	2020
Availability :	yes	**Group project :**	yes
Theme :	$\mathcal{S}$	**Framework :**	academic work

Keywords: PK-PD, Monolix
Main results:
Pharmacodynamic study of the monoclonal antibody atezolizumab for the treatment of adult patients with advanced urothelial carcinoma. Use of Monolix software to model the evolution of atezolizumab (PK-PD) concentration.

12 [$\mathcal{S}$] Effect of Methotrexate on disease free survival in patients treated for hematologic cancer with bone marrow transplants

Project size :	small	**Year :**	2020
Availability :	yes	**Group project :**	no
Theme :	$\mathcal{S}$	**Framework :**	academic work

Keywords: Cox modeling, R
Main results:
Creation of a Cox model to predict Disease Free Survival (DFS) in the case of patients treated for hematologic cancer with bone marrow transplants . Presentation of a nomogram to predict the chances of recurrence.

13 [$\mathcal{S}$] Occupational exposure to pesticides and risk of endometriosis

Project size :	small	**Year :**	2020
Availability :	yes	**Group project :**	yes
Theme :	$\mathcal{S}$	**Framework :**	academic work

Keywords: Study planning
Main results:
This project involves the development of a detailed protocol to address the general question: Does occupational exposure to pesticides increase the risk of endometriosis? To answer this question, we proposed to conduct a case-control study nested within an existing cohort.

14 [$\mathcal{B}$,$\mathcal{S}$,$\mathcal{L}$] Prediction of censorship and events from omics data (SNP) for patients treated with chemotherapy

Project size :	small	**Year :**	2020
Availability :	yes	**Group project :**	yes
Theme :	$\mathcal{B}$,$\mathcal{S}$,$\mathcal{L}$	**Framework :**	academic work

Keywords: Elasticnet, SVM, R
Main results:
In order to predict censoring or events in OMICS data, a combination of penalized regressions (Elasticnet) was used to select variables. The use of Support Vector Machine (SVM) allowed us to obtain an accuracy in our event predictions of around 70%.

15 [$\mathcal{S}$,$\mathcal{M}$] Forecasting death headcounts in metropolitan France using Bootstrap methods

Project size :	medium	**Year :**	2020
Availability :	yes	**Group project :**	no
Theme :	$\mathcal{S}$,$\mathcal{M}$	**Framework :**	academic work

Keywords: Bootstrap, Time series, R
Main results:
The objective of this project is to present different ways of forecasting the monthly number of deaths in metropolitan France and to compare these forecasts with the provisional deaths issued by INSEE. This work is therefore intended as an alternative to the traditional SARIMA models. The predictions are slightly better than those obtained using a SARIMA model.

16 [$\mathcal{S}$] Study of the French effective reproduction number R_t and its evolution during the 1^{st} quarantine period.

Project size :	medium	**Year :**	2020
Availability :	yes	**Group project :**	yes
Theme :	$\mathcal{S}$	**Framework :**	academic work

Main results:
Keywords: SIR, estimations, predictions, R
Epidemiological study on covid19 in France based on a SIR model. Estimations of the reproduction rate R_t from French data, and evidence of a decrease in this rate during quarantine ($R_0 < 1$). Estimation of the number of individuals saved thanks to quarantine from this simplistic model.

17 [$\mathcal{M}$] M/M/∞ service model applied to population dynamics as an alternative to conventional prey-predation models

Project size :	medium	**Year :**	2020
Availability :	yes	**Group project :**	yes
Theme :	$\mathcal{M}$	**Framework :**	academic work

Keywords: Birth-Death Process, Markov Chains, Ornstein-Uhlenbeck
Main results:
Use of specific Birth-Death Markov chains traditionnaly considered in queueing theories. Modeling of an ecosystem and highlighting of a stationary Poisson-like distribution. Introduction to Ornstein-Uhlenbeck processes in the context of heavy traffic. Estimation via Euler-Maruyama's method (R) in order to generate trajectories.

18 [$\mathcal{S}$] Study of antibiotic prescription times for three common diseases

Project size :	small	**Year :**	2019
Availability :	confidential	**Group project :**	no
Theme :	$\mathcal{S}$	**Framework :**	personal work

Keywords: Prescription time, R
Main results:
Participation in the production of statistics in the framework of a physician thesis. Highlighting correlations between prescription times and physician profiles (age, sex, *etc.*).

19 [$\mathcal{C}$] Creation of a Health application

Project size :	substantial work	**Year :**	2019
Availability :	partial	**Group project :**	yes
Theme :	$\mathcal{C}$	**Framework :**	academic work

Keywords: Python, Garmin, App
Main results:
A health application has been implemented in Python in order to offer a user a panel of information on his health status. Data collected via a Garmin watch and retrieved from the eponymous API.

20 [$\mathcal{G}$] Cartographic representation of the French legislative elections of 2018

Project size :	small	**Year :**	2019
Availability :	partial	**Group project :**	no
Theme :	$\mathcal{C}$	**Framework :**	personal work

Keywords: cartography, R
Main results:
The distribution of votes allocated to the various French political parties competing in 2018 is heterogeneously distributed over the french territory. This representation also makes it possible to hypothesize about the electoral base of each political party.

21 [$\mathcal{C}$] Inventory management using Python

Project size :	medium	**Year :**	2019
Availability :	yes	**Group project :**	yes
Theme :	$\mathcal{C}$	**Framework :**	academic work

Keywords: Python, SQL
Main results:
Project to create a database managed via SQL/Python to store food-related data. Implementation of a user-machine interface in Python in order to propose to a user to consult a certain amount of information related to a particular product.

22 [$\mathcal{E}$] The impact of advertising on consumption and on macroeconomic variables

Project size :	medium	**Year :**	2018
Availability :	yes	**Group project :**	yes
Theme :	$\mathcal{E}$	**Framework :**	academic work

Keywords: Bertrand assumptions, optimization
Main results:
This work is based on Bertrand's competition model. We establish a mathematical equation based on a simplistic model allowing a company in a monopolistic (even ephemeral) situation to adjust its advertising investments in an optimal way.

23 [$\mathcal{S}$,$\mathcal{E}$] Science level of 15-year-old French students: the blatant manifestation of social reproduction through extracurricular practices

Project size :	medium	**Year :**	2018
Availability :	yes	**Group project :**	yes
Theme :	$\mathcal{S}$,$\mathcal{E}$	**Framework :**	academic work

Keywords: R, SAS, multivariate exploratory statistics
Main results:
A univariate (with SAS) and then multivariate (R) study was conducted to investigate a data set from a PISA survey. There is evidence that the academic level in science of 15-year-old French students is highly correlated with the extra-curricular activities they engage in. However, the main cause of this finding is the social and cultural background of the student, which is shown here as a structure effect.

24 [$\mathcal{P}$] Modeling of heating in the frame of bicycle brakes

Project size :	substantial work	**Year :**	2017-2018
Availability :	yes	**Group project :**	yes
Theme :	$\mathcal{P}$	**Framework :**	academic work

Keywords: Python, Exponential distribution, experiments
Main results:
It has been shown that the heating of disc and pad brakes is exponentially distributed in time (from regressions, $R^2 > 95\%$). An electrical model modeling the thermodynamic heating measured during a test bench experiment allows us to describe this exponential evolution. Calorimetry experiments also permit us to obtain an estimation of some parameters of the model.

Last edited in May 2021