

TITLE

Data Scientist

Identification of biomarkers predictive of drug response by applying Machine Learning algorithms and developing R Shiny app

MISSION AND OBJECTIVES

Do you have a degree in statistics, mathematics, or a related field?

Do you want to increase your skills in Machine Learning, R, Shiny and high-dimensional data analysis?

Are you interested in expanding your knowledge in biomarkers and clinical drug discovery process?

Then join a young and dynamic international team of Data Scientists and Statisticians with our V.I.E. position in the Biomarker Statistics Team in US!

You will be involved in the analysis of thousands of genes, proteins and mutations generated in clinical trials (Oncology, Immunology and Inflammation) to understand drug response, drug action or disease progression.

You will work in close collaboration with Biomarker Statistics team members in US and Europe.

Your main missions will be:

- Explore a large catalog of Machine Learning methods (including Random Forest, Gradient Boosting, Support Vector Machine) to identify biomarkers predictive of drug response.
- Investigate deeply cross-validation for hyperparameter tuning and generalization error evaluation.
- Compare the performance of the different Machine Learning algorithms on simulated data and on Sanofi internal biomarker data.
- Evaluate the impact of unbalanced classes on performance of Machine Learning algorithms.
- Develop an R Shiny application to interactively apply different Machine Learning algorithms with different sets of hyperparameters and to evaluate and visualize performance using different quality metrics
- Promote the visibility of the R Shiny app through scientific publications, seminars, and webinars

PROFILE

- MSc in Biostatistics or relevant disciplines, e.g. Mathematics, Computer Science, Statistics (ENSAI, ISUP, ENSAE, INSA, Master 2 in Biostatistics...)
- At least 1 year of experience R programming language
- Good knowledge of R and statistics
- Knowledge of R Shiny would benefit
- Abilities : autonomous, rigorous, team player

INFORMATION

Duration: 12 months-18 months

Start date: July-September 2020

Location: Framingham MA in US

Department/Team: Biostatistics and Programming/Biomarker Statistics Team

Contact in France: caroline.paccard@sanofi.com

Link to apply:

https://sanofi.wd3.myworkdayjobs.com/SanofiCareers/job/Framingham-MA/Data-Scientist---VIE-Contract--W-M-_R2538270-1