

William Farin

Biostatisticien

Après 10 ans d'expérience en biostatistique appliquée à la recherche pharmaceutique, je me suis formé à la data science afin de combiner rigueur statistique et puissance des outils d'apprentissage automatique.



[Linkedin.com/in/william-farin](https://www.linkedin.com/in/william-farin)

[Github.com/wfarin](https://github.com/wfarin)



CONTACT



0683578655



william.farin@gmail.com



77000 La Rochette



38 ans

FORMATIONS

Formation Data Scientist

DatascienceTest - mars 2025 / juin 2025

Projet de classification multiclassée cellulaire avec réseaux de neurones de convolution

Master En bioinformatique et Biostatistique

Université Paris-Sud, Orsay, île-de-France - 2013 / 2015

Master of Bio-engineering

University of Oklahoma, Norman, USA - 2009 / 2011

Diplome d'ingénieur en genie Biologique

Polytech'Clermont-Ferrand, Clermont-Ferrand, France - 2007 / 2011

Diplome Universitaire de technologie

Université paul Sabatier, Auch, France - 2005 / 2007



EXPÉRIENCES PROFESSIONNELLES

Biostatisticien

Enterome, Paris 75000 - 2015 / 2024

-Microbiote et chirurgie bariatrique : Analyse de l'évolution du microbiote intestinal avant/après chirurgie (RYGB/SLEEVE) via shotgun metagenomics. Résultats publiés (PMID: 32360114).

-Modélisation statistique (Random Forest, régression logistique) pour prédire les stades de la NASH à partir du microbiote fécal, en alternative non-invasive à la biopsie.

-Détection de biomarqueurs microbiens fécaux prédictifs de réponse au nivolumab chez des patients atteints de cancers (gastrique, NSCLC) en Phase II, via Lasso et Random Forest pour identifier des signatures microbiennes associées à la réponse au traitement chez des patients atteints de colite.

-Analyse avancée (joint models, survie) de vaccins microbiens mimant des épitopes tumoraux (BIRC5, FOXM1) chez des patients atteints de glioblastome.

-Études RNA-Seq sur modèles rongeurs (TNBS), humaines et porcines. Développement d'un pipeline d'analyse complet (normalisation, DESeq2, enrichissement fonctionnel via GO, KEGG).

Biologiste (Master Thesis)

University of Oklahoma, Norman, USA - 2009 / 2011

Construction et évaluation d'une protéine de fusion pour cibler les adipocytes dans le cadre du développement d'une thérapie anti-obésité

R PACKAGE DEVELOPMENT

Développement et Structure

[usethis](#), [devtools](#), [roxygen2](#), [desc](#), [projroot](#), [golem](#)

Test et Qualité

[testthat](#), [covr](#), [lintr](#), [goodpratic](#)

Documentation et Site Web

[pkgdown](#), [rmarkdown](#), [knitr](#)

Publication et Vérification

[checkmate](#), [rcmdcheck](#), [rhub](#)

LANGUES

Anglais

●●●●●●●●○○○



COMPÉTENCES DE PROGRAMMATION

Méthodes Statistiques (R & Python)

●●●●●●●●●●●●

Visualisation (R shiny), PCA, Clustering, Séries Temporelles, Analyse Différentielle

Machine Learning 2 - Avancé (Python)

●●●●●●●●●●●●

Deep Learning: CNN, transformers, RNN (Keras, Pytorch & Tensorflow)

AutoML avec TPOT

Text Mining (bibliothèque re) & Webscraping

Machine Learning 1 - Classification et Regression (R & Python)

●●●●●●●●●●●●

Arbres de décision (Boosting/Strapping/RandomForest/xgboost), KNN, régression logistique, Détection d'anomalie (DBSCAN), SVM Régression Linéaire simple et multiple, Modèle Linéaire Mixte à effet aléatoires, Modèle de survie

Data Engineering/MLOps

●●●●●●●●●●●●

SQL, API, MLflow

Outils de Programmation

●●●●●●●●●●●●

GitHub, Docker, VScode, Rstudio

Outils de Présentation

●●●●●●●●●●●●

R: rmarkdown, quarto, bookdown, {officer}, flextable, DT

Python: jupyter notebook, streamlit