Olivier Binette

Statistics PhD Candidate, Duke University

github.com/OlivierBinette

olivier.binette@duke.edu

SUMMARY OF QUALIFICATIONS

- Advanced data science, machine learning, and applied research skills demonstrated through open-source projects, publications in top journals and several awards.
- Programming Skills: Python, R, C/C++, Javascript, SQL, bash, git, Linux, Docker, AWS, scikit-learn, Keras, developed through published software packages, internships, and academic research.
- Project management and leadership experience through leading interns teams for data science projects, leading large research collaborations, and managing data-labeling staff.

EXPERIENCE

Duke University **2019 – present**

Graduate Researcher

- Developed model robustness and diagnostics tools to evaluate the reliability of statistical methods for the quantification of modern slavery, resulting in a publication in the Journal of the Royal Statistical Society Series A.
- Developed entity resolution methods and software for big data integration, including the development of statistical evaluation methodology and flexible machine learning models for unsupervised and semi-supervised entity resolution.
- Taught weekly labs for *Entity Resolution*, *Spatio-Temporal Models*, *Bayesian and Modern Statistics*, and *Introduction to Data Science*.

American Institutes for Research **May 2022 – August 2022**

Data Scientist Intern

- Developed and implemented a model evaluation and improvement strategy for machine learning-based entity resolution systems used at PatentsView.org.
- Developed data labeling methodology and managed five staff labeling data for model evaluation and training.
- Led a large academic collaboration resulting in a scientific paper on estimating the performance of entity resolution systems (*arxiv.org/abs/2210.01230*) and an open-source Python package for the evaluation of entity resolution systems (*github.com/PatentsView/PatentsView-Evaluation*).

Intact Financial Corporation **January 2022 – April 2022**

Data Scientist Intern

- Optimized product line value through the development and implementation of uncertainty quantification and Bayesian optimization methods for pricing optimization in two internal Python packages.
- Collaborated with multiple stakeholders to establish new software development practices for data science tooling, addressing a long-standing lack of documentation through an integrated testing and documentation workflow.

Duke Community Food Pantry **2021 - 2022**

Research Coordinator

- Developed survey methodology for food insecurity monitoring adopted by Duke University and deployed in 2022, demonstrating widespread food insecurity in the graduate and undergraduate populations.

Information Initiative at Duke **2020 - 2022**

Project Lead

- Trained and led undergraduate students in internship projects: produced an R Shiny web app for UC Davis social sciences research group, and used remote sensing to analyze the impact of urban land use on river metabolism.

Université du Québec à Montréal **2017 – 2019**

Graduate Researcher

- Published research papers on Bayesian nonparametric inference in the Journal of Machine Learning Research, IEEE Transactions on Information Theory, and Journal of Statistical Planning and Inference.

SOFTWARE PROJECTS

StringCompare - funded by G-Research and Github Sponsors (Python, C++)
Efficient String Comparison Functions and Fuzzy String Matching.
PatentsView/PatentsView-Evaluation - funded by American Institutes for Research (Python)
Evaluation and benchmarking of PatentsView disambiguation algorithms.
dgaFast (R, C++)
Multiple Systems Estimation Using Decomposable Graphical Models in C++.
csv-search (Docker, Javascript, elasticsearch)
Quickly setup elasticsearch and a web search UI for arbitrary csv tables.
fingermatchR (R, C++)
Fingerprint matching tools based on NIST's mindtct and bozorth3 algorithms.
TessTools (R)
Tools for the use of Tesseract OCR in R.
cache - published on CRAN (R)
Easily cache and retrieve computation results in R.
assert - published on CRAN (R)
Lightweight validation tool for checking function arguments and data analysis scripts.
Fractals (Javascript, HTML)
A Javascript Mandelbrot set explorer.
Earthquakes (Javascript HTML)
3D data visualization with WebGL/three.js.

AWARDS

- G-Research PhD Student Grant - Open Source Software for Big Data Integration (2022; 2000 £)
- American Statistical Association Best Paper Award (2022)
- Canada Governor General's Academic Gold Medal (2020)
- Alexander-Graham-Bell Canada Graduate Scholarship (2019; 105 000 $)
- Fonds de Recherche du Québec - Nature et Technologies Doctoral Award (2019; 84 000 $)
- Stanford University fully-funded PhD admission offer (2019)
- Faculty of Arts and Science Top Doctoral Award (University of Toronto, 2019)
- Natural Sciences and Engineering Research Council of Canada Masters Award (2017; 21 000 $)
- Fonds de Recherche du Québec - Nature et Technologies Masters Award (2017; 21 000 $)

EDUCATION

Duke University **2019 – 2023 (expected)**
PhD Candidate, Statistical Science Department (3.9 GPA) Advisor: Prof. Jerry Reiter *Durham, NC*

Université du Québec à Montréal **2014 – 2019**
BSc, Mathematics (3.97 GPA); MSc, Statistics (4.0 GPA) *Montréal, QC*