# Bastien Dussap

★ 17 June 1998
□ 07 77 99 21 57
> bastien.dussap@inria.fr
③ bastiendussap.github.io
④ BastienDussap
○ BastienDussap
♦ bastiendussapapb

PhD student in mathematics

# Education

2021 - 2024	PhD thesis in Machine Learning, Université Paris Saclay
	PhD thesis in Machine Learning apply to the comparison of Cytometric data.

- 2019–2021 Master Mathématiques et Applications, Université Paris Saclay Master's degree in Mathematics apply to Artificial Intelligence.
- 2016–2019 Licence de Mathématiques, Université Paris Saclay Bachelor's degree of Mathematics. The first two years at Evry and the last one at Orsay.

# PhD thesis

Title Cytometric data comparison

Supervisors Gilles Blanchard and Marc Glisse

Abstract PhD thesis in partnership with Metafora biosystem a bio-engineering compary. The company has created a software, Metaflow, that allows automatic analysis of flow cytometry data. My work focuses on the use of machine learning models to transfer the analysis performed on one sample to a new unanalysed one. We rely on Reproducing Kernel Hilbert space to embed and store high-dimensional features in Euclidian Space. We use these representations to, firstly, estimate the proportions of each population in a new sample, and secondly to automatically name the cluster obtained by metaflow.

### Experience

#### Teaching

2022–2023 Mathematics for Management, *IUT Sceaux*, L1 B.U.T GEA Taught by Patrick Pamphile.

#### Seminar

2022–2024 **Seminar**, Université Paris-Saclay, Master 2 Co-organizer of the seminar for master students in Statistics and Machine Learning at Université Paris-Saclay.

# Publication

2023 Label Shift Quantification with Robust Guarantees via Distribution Feature Matching, ECML/PKDD 2023, With Gilles Blanchard and Badr-Eddine Chérief-Abdellatif

Quantification learning deals with the task of estimating the target label distribution under label shift. There exist two main classes of quantifiers in the literature: classification-based methods vs statistical mixture modeling approaches. In this paper, we propose an efficient and scalable quantifier that belongs to the second class, and we present a unifying framework based on feature distribution matching that recovers estimators from both quantification families. In particular, we derive a general consistency theorem under label shift which improves upon the bounds that can be found in the literature, investigate the misspecified setting where the exact label shift hypothesis is challenged, and provide a detailed numerical study on simulated and real-world datasets.

## Award

2023 Best Student Paper - Research Track, ECML/PKDD 2023, Label Shift Quantification with Robust Guarantees via Distribution Feature Matching

CEFRL rating

#### Languages

French Native

English B2

# Computer skills

basic knowledge	<b>EXTENSIVE</b> project experience
intermediate knowledge with some	deepened expert knowledge
project experience	expert / specialist

	Level	Skill	Years	Comment
Language:		Python	4	Used Python and standard Machine Learning packages such as numpy, matplotlib, scikit-learn or pytorch. Creation of custom packages.
		$\operatorname{SQL}$		Online course on SQLite
		R	1	Used $R$ for Machine Learning project during my education.
		Zotero	3	Used Zotero during the PhD to manage my bibliography.
		Ŀ₽ŢĘX	5	
OS:		Linux	4	I only use Ubuntu for work.
		Windows	10 +	Use Windows at home.