Baptiste Bauvin, Ph.D.

Applied Machine Learning Research Scientist,

Coactive AI, 2025

3684 Bd LaSalle
H4G 1Z6, Montréal, QC, Canada
+1 (581) 777-3782
baptiste.bauvin.work@gmail.com
baptiste-bauvin

Published papers

Journal Conference Workshop arXiv

2023 🧰🗂️ **Invariant Causal Set Covering Machines**, *T. Godon, B. Bauvin, P. Germain, J. Corbeil, A. Drouin*, arXiv, ICML Workshop SCIS

In this work, we build on ideas from the invariant causal prediction literature to propose Invariant Causal Set Covering Machines, an extension of the classical Set Covering Machine algorithm for con- & disjunctions of binary rules that provably avoids spurious associations. We demonstrate both theoretically and empirically that our method can identify the causal parents of a variable of interest in polynomial time.

2023 🧑‍🏫 **Sample Boosting Algorithm (SamBA) - An interpretable greedy ensemble classifier based on local expertise for fat data**, *B. Bauvin, C. Capponi, F. Clerc, P. Germain, S. Koço, J. Corbeil*, UAI

In this paper, we propose a supervised binary classification framework that propagates the local knowledge acquired during the boosting iterations to the prediction function. Based on this general framework, we introduce SamBA, an interpretable greedy ensemble method designed for fat datasets with a large number of dimensions and a small number of samples.

2022 🧰 **Integrating and reporting full multi-view supervised learning experiments using SuMMIT**, *B. Bauvin, J. Corbeil, D. Benielli, S. Koço, C. Capponi*, PMLR, LIDTA 2022

SuMMIT (Supervised Multi Modal Integration Tool) is a software offering many functionalities for running, tuning, and analyzing experiments of supervised classification tasks specifically designed for multi-view data set.

2022 📰 **Toolbox for Multimodal Learn (scikit-multimodallearn)**, *D. Benielli, B. Bauvin, S. Koço, R. Huusari, C. Capponi, H. Kadri, F. Laviolette*, JMLR 23(51):1—7

scikit-multimodallearn is a Python library for multimodal supervised learning, licensed under Free BSD, and compatible with the well-known scikit-learn toolbox. This paper details the content of the library, including a specific multimodal data formatting and classification and regression algorithms. Use cases and examples are also provided.

2020 📰 **Fast greedy C-bound minimization with guarantees**, *B. Bauvin, C. Capponi, J.-F. Roy & F. Laviolette*, Mach Learn 109, 1945–1986

In this work, we address the problem of accelerating the C-bound minimization process while keeping the sparsity of the solution and without losing accuracy. We present CB-Boost, a computationally efficient classification algorithm relying on a greedy–boosting-based C-bound optimization.

Experience

Company

Jan. 2024 - **Applied AI Research Scientist**, *Thales Digital Solutions*, QC, Canada

Interpretable and Trustworthy AI for edge solutions. Explaining neural networks (CNNs and ²Transformers) for the purpose of compression for embedded applications. Focusing on a Causal-based approach and physical compression of NNs. Submitted one patent and one CVPR2025 paper, both under review.

Academic

Sept. 2017 -Nov. 2023 **Machine Learning PhD**, *GRAAL-Santé*, QC , Canada

Multi-omic machine learning and interpretable algorithms. Focus on ensemble methods, form PAC-Bayesian theory to applications to biological multi-omic tasks and multi-view framework coding.

Summer 2016 **Machine learning research internship**, *GRAAL*, QC, Canada

Developing a multi-omic binary classifier based on the SCM. Graded A.

Winter 2016 **Research project**, *École Centrale Marseille (ECM)*, Marseille, France

Analysis of two multi-view algorithms, Mumbo and MVM. Graded A.

Summer 2015 **Robotics vision research internship**, *ACIN/V4R*, Vienna, Austria

Providing a path-finding algorithm for a mobile robot, based on laser and *kinect* sensors.

Summer 2014 **Machine learning research internship**, *LIF*, Marseille, France
Building a multi-class classification algorithm based on SVMs for vocal recognition. Graded A.

Winter 2013 **Research project**, *École Centrale Marseille (ECM)*, Marseille, France
Study of Γ-free matrices and Bunneman graphs for species classification. Graded A.

2012-2013 **Transversal project**, *École Centrale Marseille (ECM)*, Marseille, France
Mesh network building for pollution sensors based on Arduino boards, shared on a website. Graded A.

Summer 2013 **Production internship**, *Belle-de-Mai Incubator*, Marseille, France
Audit, debugging update of the incubator's internal management software. Graded A.

Volunteering

2023- **Tech Manager**, *Verdun Dragon Boat Club*, Verdun, Montreal
Website management, Tech advisor, Media manager, Photograph.

2020-2023 **Volunteer**, *Harmonie Richelieu*, Verdun, Montreal
Fundraising, Archive Manager.

2014 **President & Webmaster**, *Phy-TV*, ECM, Marseille
School's audiovisual association : video editing, FX, event planning, external contracts.

2014-2015 **Active Member and Regisseur**, *Bureau des Arts*, ECM, Marseille
Regisseur during on scene events, and active member

2013-2014 **Webmaster**, *Festival Massiliades*, ECM, Marseille
Website management based on DRUPAL and HTML5/CSS/PHP/MySQL.

Other

2014-2015 **Web-designer & software engineer (Front & Back-End)** , *KSI*, ECM, Marseille
Concept and development of a web-site for a client of the junior company

Education

2015-2016 **Fundamental Computer Science Master**, *Aix-Marseille Université*, Marseille, France
Focus on : Inferential Statistics, Machine Learning, Logic and Automatons, Algorithmic, Metric Spaces, Graphs and Networks Algorithmic.

2012-2016 **General Engineering Degree**, *ECM*, Marseille, France
Engineering Degree : Fundamental Computer Science, IA, Telecommunications, Micro-controllers. Major in : Computer Science, Networks, Data Science, Android. Courses in : Quantum Physics, Photonic, Mechanical Engineering

2010-2012 **Preparatory Class**, *Lycée Faidherbe*, Lille, France
Focus on Fundamental Computer Science in excellency boarding school.

Computer Science Skills

Python ML PyTorch, CUDA, ML-Flow, Kubernetes, Hydra, Scikit-learn, HDF5, Pandas, Numpy.

Languages JAVA, C++, SQL, PYTHON, Caml, Maple, Matlab.

Platforms OS X, GNU/Linux, Windows, Arduino, Android.

Web PHP, HTML5, CSS3, JAVASCRIPT(JQuery), MySQL, Drupal.

Others Any office suite, LaTeX.

Languages

Native French

Professional English

Others German (B2), Portugese, Italian, Arabic

Hobbies

Hands-on Construction, Electronics, Hardware, Mechanics, Sewing.

Sports	Dragon Boat, Hockey, Skiing, Dancing.
Cultural	Travelling (China, India, Egypt), Music (Tuba), Cooking.