Automated machine learning for batch and stream settings

Contract type: Internship
Level of qualifications required: Master's or equivalent
Function: Internship Research

Context

The MIMOVE team at Inria Paris undertakes research enabling next-generation mobile distributed systems and networks, from their conception and design to their runtime support, focusing on middleware and data. MIMOVE has longstanding expertise in mobile and service-oriented computing, semantic technologies, interoperability, system emergence and evolution, edge/fog computing, and mobile crowdsensing. MIMOVE works on these topics through a great number of national and international collaborations with academia and industry, which include large-scale software development of real-world systems. MIMOVE’s research results impact various application domains; MIMOVE focuses in particular on the application areas of IoT, smart cities, and citizen participation.

Inria MIMOVE is one of the 13 partners of the SEDIMARK project (https://sedimark.eu/) of the Horizon Europe Framework Programme. The EU has a vision to become the world’s largest and most secure Data Hub where businesses can exploit data for promoting research, education, and business. This requires significant investments in the development of (among others) intelligent tools for data interoperability, secure data sharing, and data discovery. SEDIMARK aims to provide an enriched secure decentralised data and services marketplace, where scattered data from various domains and geographical locations within the EU can be easily generated, cleaned, protected, discovered, enriched with metadata, AI and analytics and exploited for diverse business and research scenarios.

This internship is part of the SEDIMARK project. The selected candidate will be supervised by Maroua Bahri (maroua.bahri@inria.fr) and Nikolaos Georgantas (nikolaos.georgantas@inria.fr) who are working on this project.

Assignment

Automated Machine Learning (AutoML) is an approach that optimizes the machine learning process, making it more accessible and efficient for users with varying levels of expertise [1]. AutoML leverages algorithms and computational power to automate key tasks in the machine learning pipeline, such as feature engineering, model selection, and hyperparameter tuning. This enables individuals without extensive machine learning knowledge to build and deploy effective models. This project will focus on autoML with a specific emphasis on hyperparameter optimization for both supervised and unsupervised learning, tailored for batch and real-time data stream settings [2]. SEDIMARK aims at employing autoML to automate various tasks and optimize the overall data pipeline as a whole.

Main activities

This internship aims to contribute to the evolution of AutoML frameworks by enhancing their adaptability and efficiency. This will involve exploring and integrating algorithms into the AutoML pipeline, with a focus on optimizing hyperparameters to enhance model performance and selection. Additionally, the project will address the challenges posed by data streams, aiming to develop innovative approaches for hyperparameter optimization that consider the dynamic nature of streaming data [3][4]. The expected outcomes include an improved AutoML framework capable of efficient and effective hyperparameter optimization and insights for deploying AutoML models in different scenarios.

References:


Skills

Offer #2023-07000
- Master level research internship (M2) or equivalent (stage de fin d'études ingénieur)
- Expertise in Python programming
- Sound knowledge of machine learning
- Knowledge of optimization concepts

Benefits package

- Subsidized meals
- Partial reimbursement of public transport costs
- Leave: 7 weeks of annual leave + 10 extra days off due to RTT (statutory reduction in working hours) + possibility of exceptional leave (sick children, moving home, etc.)
- Professional equipment available (videoconferencing, loan of computer equipment, etc.)
- Social, cultural and sports events and activities
- Access to vocational training
- Social security coverage

General Information

- Theme/Domain: Data and Knowledge Representation and Processing
  Data production, processing, analysis (BAP D)
- Town/city: Paris
- Inria Center: Centre Inria de Paris
- Starting date: 2024-03-01
- Duration of contract: 6 months
- Deadline to apply: 2024-01-31

Contacts

- Inria Team: MIMOVE
- Recruiter: Bahri Maroua / maroua.bahri@inria.fr

About Inria

Inria is the French national research institute dedicated to digital science and technology. It employs 2,600 people. Its 200 agile project teams, generally run jointly with academic partners, include more than 3,500 scientists and engineers working to meet the challenges of digital technology, often at the interface with other disciplines. The Institute also employs numerous talents in over forty different professions. 900 research support staff contribute to the preparation and development of scientific and entrepreneurial projects that have a worldwide impact.

Warning: you must enter your e-mail address in order to save your application to Inria. Applications must be submitted online on the Inria website. Processing of applications sent from other channels is not guaranteed.

Instruction to apply

Defence Security:
This position is likely to be situated in a restricted area (ZRR), as defined in Decree No. 2011-1425 relating to the protection of national scientific and technical potential (PPST). Authorisation to enter an area is granted by the director of the unit, following a favourable Ministerial decision, as defined in the decree of 3 July 2012 relating to the PPST. An unfavourable Ministerial decision in respect of a position situated in a ZRR would result in the cancellation of the appointment.

Recruitment Policy:
As part of its diversity policy, all Inria positions are accessible to people with disabilities.