



Activity Report 2022

Team LACODAM

Large scale Collaborative Data Mining

Joint team with Centre Inria de l'Université de Rennes

D7 – Data and Knowledge Management



Contents

Project-Team LACODAM	1
1 Team members, visitors, external collaborators	3
2 Overall objectives	4
3 Research program	5
3.1 Introduction	5
3.2 Symbolic methods	5
3.3 Interpretable ML	6
3.4 Real world AI	6
4 Application domains	6
4.1 Industry	7
4.2 Health	7
4.3 Robotics	7
4.4 Agriculture and Environment	8
4.5 Education	8
4.6 Semantic Data Management	8
5 Social and environmental responsibility	9
5.1 Footprint of research activities	9
5.2 Impact of research results	9
6 Highlights of the year	10
6.1 Awards	10
6.2 White book	10
6.3 Life of the team	10
7 New software and platforms	10
7.1 New software	10
7.1.1 HIPAR	10
8 New results	10
8.1 Symbolic Methods	10
8.1.1 Pattern Mining	10
8.1.2 Graph-FCA	12
8.1.3 Semantic Web	12
8.2 Interpretable Machine Learning	13
8.3 Real World AI	14
8.3.1 Computer Vision and Robotics	14
8.3.2 Recommender Systems	15
8.3.3 Forecasting	15
8.3.4 Logistics	15
8.3.5 Privacy	16
8.3.6 Medicine	16
8.3.7 Agriculture	17
9 Bilateral contracts and grants with industry	18
9.1 Bilateral contracts with industry	18

10 Partnerships and cooperations	20
10.1 International research visitors	20
10.1.1 Visits of international scientists	20
10.1.2 Visits to international teams	21
10.2 European initiatives	21
10.2.1 H2020 projects	21
10.3 National initiatives	21
10.3.1 ANR	23
10.4 Regional initiatives	24
11 Dissemination	25
11.1 Promoting scientific activities	25
11.1.1 Scientific events: organisation	25
11.1.2 Scientific events: selection	25
11.1.3 Journal	26
11.1.4 Invited talks	27
11.1.5 Leadership within the scientific community	28
11.1.6 Scientific expertise	28
11.1.7 Research administration	28
11.2 Teaching - Supervision - Juries	29
11.2.1 Teaching	29
11.2.2 Supervision	30
11.2.3 Juries	31
11.3 Popularization	32
11.3.1 Education	32
11.3.2 Interventions	32
12 Scientific production	32
12.1 Major publications	32
12.2 Publications of the year	33
12.3 Cited publications	36

Project-Team LACODAM

Creation of the Project-Team: 2017 November 01

Keywords

Computer sciences and digital sciences

- A2.1.5. – Constraint programming
- A3.1.1. – Modeling, representation
- A3.1.2. – Data management, quering and storage
- A3.1.6. – Query optimization
- A3.1.11. – Structured data
- A3.2.1. – Knowledge bases
- A3.2.2. – Knowledge extraction, cleaning
- A3.2.3. – Inference
- A3.2.4. – Semantic Web
- A3.3. – Data and knowledge analysis
- A3.3.1. – On-line analytical processing
- A3.3.2. – Data mining
- A3.3.3. – Big data analysis
- A3.4.1. – Supervised learning
- A3.4.2. – Unsupervised learning
- A3.4.3. – Reinforcement learning
- A3.4.4. – Optimization and learning
- A3.4.5. – Bayesian methods
- A3.4.6. – Neural networks
- A3.4.8. – Deep learning
- A3.5.2. – Recommendation systems
- A4.7. – Access control
- A5.1. – Human-Computer Interaction
- A5.2. – Data visualization
- A5.3. – Image processing and analysis
- A5.3.2. – Sparse modeling and image representation
- A5.4.1. – Object recognition
- A5.4.6. – Object localization
- A5.4.7. – Visual servoing
- A9.1. – Knowledge
- A9.2. – Machine learning
- A9.3. – Signal analysis

A9.4. – Natural language processing

A9.6. – Decision support

A9.7. – AI algorithmics

A9.8. – Reasoning

A9.10. – Hybrid approaches for AI

Other research topics and application domains

B3.5. – Agronomy

B3.6. – Ecology

B3.6.1. – Biodiversity

B9.1. – Education

B9.5.6. – Data science

1 Team members, visitors, external collaborators

Research Scientist

- Luis Galarraga Del Prado [INRIA]

Faculty Members

- Alexandre Termier [Team leader, UNIV RENNES I, Professor, HDR]
- Tassadit Bouadi [UNIV RENNES I, Associate Professor]
- Peggy Cellier [INSA RENNES, Associate Professor, HDR]
- Sebastien Ferre [UNIV RENNES I, Professor, from Apr 2022, HDR]
- Elisa Fromont [UNIV RENNES I, Professor, HDR]
- Romaric Gaudel [UNIV RENNES I, Associate Professor, from Sep 2022]
- Christine Largouet [L'INSTITUT AGRO, Associate Professor, HDR]
- Véronique Masson [UNIV RENNES I, Associate Professor]
- Laurence Rozé [INSA RENNES, Associate Professor]

PhD Students

- Abderaouf Nassim Amalou [UNIV RENNES I, with PACAP Team]
- Hugo Ayats [UNIV RENNES I, from Apr 2022]
- Johanne Bakalara [UNIV RENNES I, ATER, until Aug 2022]
- Francesco Bariatti [UNIV RENNES I, ATER, until Aug 2022]
- Julie Boudebs [UNIV RENNES I, from Apr 2022]
- Simon Corbille [UNIV RENNES I, with INTUIDOC Team]
- Lenaig Cornanguer [INRIA]
- Julien Delaunay [INRIA]
- Samuel Felton [UNIV RENNES I, with RAINBOW Team]
- Olivier Gauriau [ACTA, CIFRE]
- Camille-Sovanneary Gauthier [LOUIS VITTON, CIFRE, until Mar 2022]
- Elodie Germani [UNIV RENNES I, with EMPENN Team]
- Victor Guyomard [ORANGE LABS, CIFRE]
- Gwladys Kelodjou [UNIV RENNES I, from Nov 2022]
- Lucie Lepetit [INRIA, from Oct 2022]
- Grégory Martin [STELLANTIS, CIFRE, until Sep 2022]
- Pierre Maurand [INSA RENNES, from Oct 2022]
- Josie Signé [INRIA, until Mar 2022]
- Antonin Voyez [ENEDIS, CIFRE, with SPICY Team]

Technical Staff

- Louis Bonneau De Beaufort [L'INSTITUT AGRO, Engineer, Research Engineer]

Interns and Apprentices

- Sarah Ababou [INRIA, Intern, from Mar 2022 until Sep 2022]
- Mensah-David Assigbi [INRIA, Intern, from May 2022 until Jul 2022]
- Thomas Bobille [UNIV RENNES I, Intern, from Oct 2022]
- Arnauld-Cyriaque Djedjemel [INRIA, Intern, from May 2022 until Aug 2022]
- Cindy Ferreira [INRIA, Intern, from Mar 2022 until Aug 2022, with ORPAILLEUR Team]
- Anas Katim [INRIA, Intern, from Jun 2022 until Aug 2022]
- Lucie Lepetit [INRIA, Intern, from Mar 2022 until Aug 2022]
- Pierre Maurand [INRIA, Intern, from Mar 2022 until Aug 2022]
- Pierre Nunn [ENS Rennes, Intern, from May 2022 until Jul 2022]
- Paul Sevellec [INRIA, Intern, from May 2022 until Aug 2022]
- François Wallyn [UNIV RENNES 1, Intern, from Jun 2022 until Jun 2022]

Administrative Assistant

- Gaëlle Tworkowski [INRIA]

Visiting Scientists

- Assie Brou Ida [UFHB, Abidjan, Côte d'Ivoire, from May 2022 until Jun 2022]
- Bruno Crémilleux [Univ CAEN, from Mar 2022 until Jul 2022, 5 months Delegation Cnrs]
- Gonzalo Mendez [ESPOL, Equateur, from Mar 2022, two "one month" visits]

2 Overall objectives

Data collection is ubiquitous nowadays and it is providing our society with tremendous volumes of knowledge about human, environmental, and industrial activity. This ever-increasing stream of data holds the keys to new discoveries, both in industrial and scientific domains. However, those keys will only be accessible to those who can make sense out of such data. This is, however, a hard problem. It requires a good understanding of the data at hand, proficiency with the available analysis tools and methods, and good deductive skills. All these skills have been grouped under the umbrella term “Data Science” and universities have put a lot of effort in producing professionals in this field. “Data Scientist” is currently an extremely sought-after job, as the demand far exceeds the number of competent professionals. Despite its boom, data science is still mostly a “manual” process: current data analysis tools still require a significant amount of human effort and know-how. This makes data analysis a lengthy and error-prone process. This is true even for data science experts, and current approaches are mostly out of reach of non-specialists.

The objective of the team LACODAM is to facilitate the process of making sense out of (large) amounts of data. This can serve the purpose of deriving knowledge and insights for better decision-making. Our approaches are mostly dedicated to provide novel tools to data scientists, that can either perform tasks not addressed by any other tools, or that improve the performance in some area for existing tasks (for instance reducing execution time, improving accuracy or better handling imbalanced data).

3 Research program

3.1 Introduction

LACODAM is a research team on data science methods and applications, composed of researchers with a background in symbolic AI, data mining, databases, and machine learning. Our research is organized along the three following research axes:

- **Symbolic methods** (Section 3.2) is the first fundamental research axis. It focuses on methods that operate in symbolic domains, that usually take as input discrete data (ex: event logs, transactional data, RDF data) and output symbolic results (ex: patterns, concepts).
- **Interpretable Machine Learning** (Section 3.3) is the other fundamental research axis of the team. It aims at providing interpretable machine learning approaches, mostly by proposing *post-hoc interpretability* for state-of-the-art numerical machine learning methods. *Interpretable by design* machine learning approaches that do not fall into the "Symbolic methods" axis are also studied here.
- **Real world AI** (Section 3.4) deals with the application or adaptation of the methods developed in the aforementioned fundamental axes to real world problems. These works are conducted in collaboration with either industrial or academic partners from other domains. For example, one important application area for the team is numerical agriculture with colleagues from Inrae.

3.2 Symbolic methods

LACODAM's core symbolic expertise is in methods for exploring efficiently large combinatorial spaces. Such expertise is used in three main research areas:

- Pattern mining, a field of data mining where the goal is to find regularities in data (in an unsupervised way);
- Semantic web, where the goal is to reason over the contents of the Web;
- Skyline queries, where the goal is to find solutions to multiple criteria optimization queries.

In the pattern mining domain, the team is well known for tackling problems where the data and expected patterns have a temporal components. Usually the data considered are timestamped event logs, an ubiquitous type of data nowadays. The patterns extracted can be more or less complex subsequences, but also patterns exhibiting temporal periodicity.

A well-known problem in pattern mining is pattern explosion: due to either underspecified constraints or the combinatorial nature of the search space, pattern mining approaches may produce millions of patterns of mixed interest. The current best approach to limit the number of output patterns is to produce a small size *pattern set*, where the set optimizes some quality criteria. The best pattern set methods so far are based on information theory and rely on the principle of Minimum Description Length (MDL). LACODAM is the leading French team on MDL-based pattern mining, especially for complex patterns. After having integrated Peggy Cellier in 2021, who is the main French expert in MDL-based pattern mining, we integrated this year Sébastien Ferré, who is also an expert in this area, especially for graph patterns.

The contribution of the team in the Semantic Web domain focuses on different problems related to knowledge graphs (KGs) – usually extracted (semi-)automatically from the Web. These include applications such as mining and reasoning, as well as data management tasks such as provenance and archiving. Reasoning can resort to either symbolic methods such as Horn rules or numeric approaches such as KG embeddings that can be explained via post-hoc explainability modules. The integration of Sébastien Ferré (former SemLIS team leader) further strenghtens the Semantic Web axis by extending our expertise on general graph mining, relation extraction, and semantic data exploration.

Skyline queries is a research topic from the database community, and is closely related to multi-criteria optimization. In transactional data, one may want to optimize over several different attributes of equal importance, which means discovering a Pareto Front (the "skyline"). The team has expertise

on skyline queries in traditional databases as well as their application to pattern mining (extraction of *skypatterns*). Recently, the team started to tackle the extraction of skyline *groups*, i.e. groups of records that together optimize multiple criteria.

3.3 Interpretable ML

Making Machine Learning more interpretable is one of the greatest challenges for the AI community nowadays. LACODAM contributes to the main areas of explainable AI (XAI):

- From a fundamental point of view, the team is trying to deepen the understanding of state-of-the-art post-hoc interpretability approaches (LIME/SHAP), in order to improve these methods or adapt them to novel domains. The team has also started working on the generation of counterfactual explanations. Both lines of work have in common the need for novel notions of neighborhood of points in the model's data space.
- The team is also working on “interpretable-by-design” machine learning methods, where the decision taken can immediately be explained by the (part of) the model that took the decision. Approaches used can as well be deep learning architectures or hybrid numeric/symbolic models relying on pattern mining techniques.
- Last, the team has a special interest in time series data, which arises in many applications but has not yet received enough attention from the interpretability community. We have proposed both post-hoc and “by design” approaches for interpretable ML for time series.

More generally, LACODAM is interested in the study of the interpretability-accuracy trade-off. Our studies may be able to answer questions such as “how much accuracy can a model lose (or perhaps gain) by becoming more interpretable?”. Such a goal requires us to define interpretability in a more principled way—a challenge that has very recently been addressed, not yet overcome.

3.4 Real world AI

LACODAM's research work is firmly rooted in applications. On the one hand the data science tools proposed in our fundamental work need to prove their value at solving actual problems. And on the other hand, working with practitioners allows us to understand better their needs and the limitations of existing approaches w.r.t. those needs. This can open new and fruitful (fundamental) research directions.

Our objective, in that axis, is to work on challenging problems with interesting and pertinent partners. We target problems where off-the-shelf data science approaches either cannot be applied or do not give satisfactory results: such problems are the most likely to lead to new and meaningful research in our field. For some problems, collaborative research may not necessarily lead to fundamental breakthroughs, but can still allow making progress in the practitioners' field. We also value such work, which contributes to the discovery of new knowledge and helps industrial partners innovate.

Due to the team expertise in handling temporal data, a lot of our applicative collaborations revolve around the analysis of time series or event logs. Naturally, our work on interpretability is also present in most of our collaborations, as experts want accurate models, but also want to understand the decisions of those models.

The precise application domains are described in more details in the next section (Section 4).

4 Application domains

The current period is extremely favorable for teams working in Data Science and Artificial Intelligence, and LACODAM is not the exception. We are eager to see our work applied in real world applications, and have thus an important activity in maintaining strong ties with industrial partners concerned with marketing and energy as well as public partners working on health, agriculture and environment.

4.1 Industry

We present below our industrial collaborations. Some are well established partnerships, while others are more recent collaborations with local industries that wish to reinforce their Data Science R&D with us.

- **Car Sharing Data Analysis.** Peugeot-Citroën (PSA) group's know-how encompasses all areas of the automotive industry, from production to distribution and services. Among others, its aim is to provide a car sharing service in many large cities. This service consists in providing a fleet of cars and a “free floating” system that allows users to use a vehicle, then drop it off at their convenience in the city. To optimize their fleet and the availability of the cars throughout the city, PSA needs to analyze the trajectory of the cars and understand the mobility needs and behavior of their users. We have tackled this subject together through the CIFRE PhD of Gregory Martin, defended in December 2022.
- **Recommender Systems.** The CIFRE PhD of Camille-Sovanneary Gauthier at Louis Vuitton is concerned with the identification of the right click behavioral models for clients in order to optimize the arrangement of the items presented to potential customers in Web pages. This work builds upon new bandit algorithms to infer the parameters that model the customers' behavioral patterns accurately. It was defended in March 2022.
- **Privacy-Preserving Data-Sharing** The collection of electrical consumption time series through smart meters grows with ambitious nationwide smart grid programs. This data is both highly sensitive and highly valuable: strong laws about personal data protect it while laws about open data aim at making it public after a privacy-preserving data publishing process. The CIFRE PhD of Antonin Voyez, founded by Enedis, is concerned with this application. We study the uniqueness of large scale real-life fine-grained electrical consumption time-series, the potential privacy threats, and their mitigation.

4.2 Health

- **Care Sequences for the Exploration of Medico-administrative Data.** The difficulty of analyzing medico-administrative data is the semantic gap between the raw data (for example, database record about the delivery at date t of drug with ATC 2 code N 02BE01) and the nature of the events sought by clinicians (“was the patient exposed to a daily dose of paracetamol higher than 3g?”). The solution that is used by epidemiologists consists in enriching the data with new types of events that, on the one side, could be generated from raw data and on the other side, have a medical interpretation. Such new abstract events are defined by clinician using proxies. For example, drugs deliveries can be translated in periods of drug exposure (drug exposure is a time-dependent variable for non-random reasons) or identify patient stages of illness, etc. A proxy can be seen as an abstract description of a care sequence.

Currently, the clinicians are limited in the expression of these proxies both by the coarse expressivity of their tools and by the need to process efficiently large amount of data. From a semantic point of view, care sequences must fully integrate the temporal and taxonomic dimensions of the data to provide significant expression power. From a computational point of view, the methods employed must make it possible to efficiently handle large amounts of data (several millions care pathways). The aim of the PhD of Johanne Bakalara was to study temporal models of sequences in order to 1) show their abilities to specify complex proxies representing care sequences needed in pharmaco-epidemiological studies and 2) build an efficient querying tool able to exploit large amount of care pathways. The PhD was defended in June 2022.

4.3 Robotics

- **Visual Servoing.** Visual servoing (VS) is the task of controlling a robot by means of a camera, and is a common way to provide instructions to robots nowadays. The PhD thesis of Samuel Felton has for purpose the exploration of novel deep learning techniques, and unsupervised learning to improve the quality of VS settings and reduce the amount of human work to provide training data to such systems. This project is joint work with the RAINBOW team (IRISA).

4.4 Agriculture and Environment

- **Animal welfare.** There has been an increasing concern of both consumers and professionals to better take into account farm animals welfare. For consumers, this is an important ethical issue. For professionals, their animals will have to be able to adapt to quickly evolving climatic conditions due to global warming, thus required to improve animal health and resilience. Better understanding animal welfare in a key component of these improvements. This is the general topic of the WAIT4 project (see Section 10.3), where Lacodam provides its data mining expertise to analyze time series of precision farming sensors, as well as event logs of animal behaviors. As a first topic of research in this project, the PhD of Lucie Lepetit is concerned with heat stress. The data are rumen temperature data from dairy cows of our Inrae partner. In this data, we can notice that in especially hot days of summer, some cows have difficulties to cope with the high temperature and while exhibit high rumen temperature both during the event and during several days after. While on the other hand, there are cows that are only mildly affected by the heat during the event, and who will quickly resume to a normal rumen temperature. Our goal is to design a method that quickly identifies all the abnormal rumen temperature periods correlated to high external temperature, and that provides a characterization of the cows that either resist well to the heat, or on the contrary do not cope well with it.
- **Ecosystem Modeling and Management.** Ongoing research on ecosystem management includes modelling of ecosystems and anthropogenic pressures, with a special concern on the representation of socio-economical factors that impact human decisions. A main research issue is how to represent these factors and how to integrate their impact on the ecosystem simulation model. This work is an ongoing cooperation with ecologists from the Marine Spatial Ecology of Queensland University, Australia and from Agrocampus Ouest.

Prediction of the Dynamics of Crop Diseases. The PhD thesis of Olivier Gauriau focuses on the prediction of the dynamics of crop diseases by means of pattern-aided regression techniques. Such techniques are known to strike an interesting trade-off between accuracy and interpretability, which can help agronomers understand the best predictors of high disease incidence, and therefore optimize the usage of phytosanitary products. This project is funded by #DigitAg and the Ecophyto program and constitutes a collaboration with the ACTA of Toulouse and the INRAE.

4.5 Education

- **Data-oriented Academic Counseling.** Course selection and recommendation are important aspects of any academic counseling system. The Learning Analytics community has long supported these activities via automatic, data-based tools for recommendation and prediction. LACODAM, in collaboration with the Ecuadorian research center CTI¹ has contributed to this body of research with the design of a tool that allows students to select multiple courses and predict their academic performance based on historical academic data. The tool resorts to visualization and interpretable machine learning techniques, and is intended to be used by the students before the counseling sessions to plan their upcoming semester at the Ecuadorian university ESPOL. In our ongoing collaboration with CTI we are studying the impact of academic predictions, explanations in the behavior and decision of the students and counselors.
- **Online Children Handwriting Recognition.** The PhD thesis of Simon Corbillé addresses the problem of online handwriting recognition, a problem that enjoys satisfactory solutions for adults, but remains a challenge for children. This is because, children handwriting is, at an early stage of learning, approximate and includes deformed letters. This is a joint effort between the LACODAM and IntuiDoc (IRISA) teams.

4.6 Semantic Data Management

- **RDF Archiving and Provenance.** Archiving and provenance tracking are two crucial tasks in the management of large collaborative RDF knowledge bases, such as Wikidata or DBpedia. This is a

¹Centro de Tecnologías de Información

consequence of the dynamicity and source heterogeneity of such data collections. Notwithstanding the value of RDF archiving and provenance tracking for both data maintainers and consumers, this field of research remains under-developed for multiple reasons. These include, among others, the lack of usability and scalability of the existing systems, a disregard of the evolution patterns of RDF datasets, and a weaker focus on data processes involving non-monotone operations². These challenges are tackled in our ongoing collaboration with the DAISY team of Aalborg University, namely thanks the PhD thesis of Olivier Pelgrin on scalable RDF archiving, and the post-doctoral fellowship of Daniel Hernández on how-provenance computation for SPARQL queries.

5 Social and environmental responsibility

5.1 Footprint of research activities

There are two main axes that characterize the bulk of LACODAM's environmental impact: work trips, and computing resources utilisation.

Work trips. While the sanitary crisis had drastically cut the quantity of work trips of the team, the year 2022 has seen an increase in the physical participation to conferences and various committees. However compared to the pre-covid period, one can note that the majority of movements are national or at best European, with very few trips outside of Europe. This may change in 2023 with some conference acceptance in the US. It seems that in general, the possibility of participating to meetings by videoconference has removed many “low added value trips”. This is a first step in reducing our carbon footprint in a meaningful way, while preserving some of trips important for the scientific as well as human aspect of our work.

Utilisation of computing resources. LACODAM contributed in 2020 with a new server (abacus12) to the Igrida computing platform. Being a team specialized in data science and machine learning, a recurrent task in LACODAM is to run CPU intensive algorithms on large data collections, for example, to train deep neural networks. Some of our ongoing PhD research topics (e.g., the theses of Simon Corbillé, and Simon Felton) concern deep learning technologies, and the increasing place of eXplanaible AI in our research program will boost our reliance on Igrida (notably with the PhD of Julien Delaunay and Victor Guyomard). This will increase the energetic and environmental footprint of our activities in a non-negligible way. We are therefore willing to collaborate with the institute's direction in any initiative that could mitigate such an impact.

5.2 Impact of research results

We estimate that the research work can have actual impact in three different ways:

- In the short/medium term, a significant part of our research work is conducted in collaboration with companies, through CIFRE PhDs. Hence, the addressed research problems concern an important challenge for the company, and the solutions proposed are evaluated on their relevance to tackle this challenge.
- In the medium/long term, we also have potential impactful research work with scientists from other domains, especially in environment and agriculture. Some earlier work of the team, conducted with INRAE SAS team, helped better understand nitrate pollution in Brittany, an important environmental issue. Current work of Lucie Lepetit are dedicated to the design of better data mining tools to characterize heat stress for the cows, which will help to guarantee the well being of farm animals in a time of climate change.
- Last, in the longer term, the team has a fundamental line of work on machine learning and interpretability. This is a critical topic nowadays due to the emergence of the GDPR. Given the increasing use of machine learning solutions in most areas of human activity, work on interpretability is of

²Processes where there is some sort of data difference

utmost societal importance, as it will help in designing more useful and also more acceptable machine learning approaches. This will require a sustained effort from the community: LACODAM is taking part in this effort, both on its own, as the coordinator of the Inria HyAIAI project, and last by having several of its members in the large European Project TAILOR dedicated to this topic.

6 Highlights of the year

6.1 Awards

- Francesco Bariatti was awarded the EGC 2023 Thesis Prize for his PhD he did under the supervision of Sébastien Ferré.

6.2 White book

Alexandre Termier was in the group of authors of Inrae-Inria white book on digital agriculture [bellonmaurel:hal-03604970] [bellonmaurel:hal-03900024], that summarizes the common vision of the two institutes on the state of the art in this domain, and in the future challenges that will have to be addressed. In the team, Tassadit Bouadi, Véronique Masson and Christine Largouët all contributed to the writing of this book.

6.3 Life of the team

Sébastien Ferré, former head of the SemLIS team of IRISA, joined us in April 2022. With him came three PhD students of SemLIS that he co-supervised with Peggy Cellier: Francesco Bariatti (left for a postdoc in Leiden in August), Hugo Ayats and Julie Boudebs. Thanks to the proximity of the research topics and to our numerous informal exchanges, the integration went smoothly and the team is grateful for these new members.

7 New software and platforms

7.1 New software

7.1.1 HIPAR

Name: Hierarchical Interpretable Pattern-aided Regression

Keywords: Regression, Pattern extraction

Functional Description: Given a (tabular) dataset with categorical and numerical attributes, HIPAR is a Python library that can extract accurate hybrid rules that offer a trade-off between (a) interpretability, (b) accuracy, and (c) data coverage.

URL: <https://gitlab.inria.fr/lgalarra/hipar>

Contact: Luis Galarraga Del Prado

8 New results

We organize the scientific results of the research conducted at LACODAM according to the axes described in our research program (Section 3).

8.1 Symbolic Methods

8.1.1 Pattern Mining

Participants: Tassadit Bouadi, Peggy Cellier, Lénaïg Cornanguer, Christine Largouët, Laurence Rozé, Alexandre Termier.

Remark about the “Participants” boxes: we compiled syntactically the list of co-authors of the papers that make the “New Results” of the year, for each subsection. It obviously does not mean that other members of the team do not work on the topics listed, the correct meaning is that they did not have a publication on that topic this year.

TAG: Learning Timed Automata from Logs [18]. Event logs are often one of the main sources of information to understand the behavior of a system. While numerous approaches have extracted partial information from event logs, in this work, we aim at inferring a global model of a system from its event logs. We consider real-time systems, which can be modeled with Timed Automata: our approach is thus a Timed Automata learner. There is a handful of related work, however, they might require a lot of parameters or produce Timed Automata that either are undeterministic or lack precision. In contrast, our proposed approach, called TAG, requires only one parameter and learns a deterministic Timed Automaton having a good tradeoff between accuracy and complexity of the automata. This allows getting an interpretable and accurate global model of the real-time system considered. Our experiments compare our approach to the related work and demonstrate its merits.

QuickFill, QuickMixte: block approaches for reducing the number of programs in program synthesis [22]. Repetitive tasks are often tedious; in order to facilitate their execution, program synthesis approaches have been developed. They consist in automatically inferring programs that satisfy a user’s intention. The best-known approach to program synthesis is FlashFill, which is integrated into the Excel spreadsheet and allows the processing of strings. In FlashFill the user’s intention is represented by examples i.e. pairs (input, output). FlashFill explores a very large space of programs and can therefore require a large execution time and infer many programs, some of which work on given examples but do not capture the user’s intent. In this paper, we propose two block-based QuickMix and QuickFill approaches that aim to guide the exploration of FlashFill’s program space by enriching the user-supplied specifications. These approaches require the user to provide associations between output and input subparts to refine the specifications. Experiments conducted on a set of 12 datasets show that QuickMix and QuickFill significantly reduce the program space of FlashFill. We show that with these approaches, it is often possible to give fewer examples than with the original FlashFill algorithm for a higher proportion of correct programs.

On computing evidential centroid through conjunctive combination: an impossibility theorem [14]. The theory of belief functions (TBF) is now a widespread framework to deal and reason with uncertain and imprecise information, in particular to solve information fusion and clustering problems. Combination functions (rules) and distances are essential tools common to both the clustering and information fusion problems in the context of TBF, which have generated considerable literature. Distances and combination between evidence corpus of TBF are indeed often used within various clustering and classification algorithms, however their interplay and connections have seldom been investigated, which is the topic of this paper. More precisely, we focus on the problem of aggregating evidence corpus to obtain a representative one, and we show through an impossibility theorem that in this case, there is a fundamental contradiction between the use of conjunctive combination rules on the one hand, and the use of distances on the other hand. Rather than adding new methodologies, such results are instrumental in guiding the user among the many methodologies that already exist. To illustrate the interest of our results, we discuss different cases where they are at play. Impact Statement-Within the theory of belief functions, both distances and conjunctive combination rules can be used to achieve very similar purposes: evaluating the conflict between sources, performing supervised or unsupervised learning in presence of evidential information, or more simply obtaining a synthetic representation of multiple items of information. However, the results obtained by both approaches may show some inconsistency between them. This paper provides some insight as to why this may happen, showing that the two approaches are definitely at odds, and that using distances is, for instance, incompatible with

some fundamental notions of the theory of belief functions, such as the least commitment principle. We illustrate the importance of the studied differences on problems such as k-centroid clustering, and discuss the importance of interpretations in such problems, which is rarely done in the literature.

8.1.2 Graph-FCA

Participants: Hugo Ayats, Peggy Cellier, Sébastien Ferré.

CONNOR: Exploring Similarities in Graphs with Concepts of Neighbors [17]. Since its first formalization, the Formal Concept Analysis (FCA) field has shown diverse extensions of the FCA paradigm. A recent example is Graph-FCA, an extension of FCA to graphs. In the context of Graph-FCA, a notion of concept of neighbors has been introduced to support a form of nearest neighbor search over the nodes of a graph. Concepts of neighbors have been used for diverse tasks, such as knowledge graph completion and relation classification in texts. In this paper, we present CONNOR, a Java library for the computation of concepts of neighbors on RDF graphs.

Exploring the Application of Graph-FCA to the Problem of Knowledge Graph Alignment [20]. Knowledge Graphs (KG) have become a widespread knowledge representation. When different KGs exist for some domain, it is valuable to merge them into a richer KG. This is known as the problem of KG alignment, which encompasses related problems such as entity alignment or ontology matching. Although most recent approaches rely on supervised representation learning, Formal Concept Analysis (FCA) has also been proposed as a basis for symbolic and unsupervised approaches. We here explore the application of Graph-FCA, an extension of FCA for KGs, to different scenarios of KG alignments: (A) when the two KGs have common values, and (B) when pre-aligned pairs are known. We show that, compared to previous FCA-based approaches, Graph-FCA allows for a more natural and scalable representation of the KGs to be aligned, and makes it simpler to extract alignments from the concepts. It also features flexibility w.r.t. different alignment scenarios.

Modeling Complex Structures in Graph-FCA: Illustration on Natural Language Syntax [21]. Graph-FCA is an extension of formal concept analysis for multi-relational data. In this paper, we discuss the freedom of representation offered by Graph-FCA, in particular by its support of n-ary relations, considering natural language syntax as a use case.

8.1.3 Semantic Web

Participants: Hugo Ayats, Peggy Cellier, Sébastien Ferré.

Some of previously presented documents also contribute to this research domain: [17] [20] .

Construction de Graphes de Connaissance à partir de textes avec une I.A. centrée-utilisateur [27]. With the rise of the Semantic Web in the last two decades, a need for tools to build good quality knowledge graphs has emerged. This paper presents the design of an explainable, user-centered method for the semi-automated production of knowledge graphs from domain-specific texts. This system is initially presented as a guided RDF editing interface. Then, based on the user's actions, a triplet suggestion system is implemented. Finally, through interactions with the user, the system gradually automates the process. After presenting the workflow of the system and detailing the units that compose it - a pre-processing unit, an interactive unit and an automated unit - this article documents the aspects of this workflow that have already been implemented, as well as the results of their evaluation.

A Two-Step Approach for Explainable Relation Extraction [16]. Knowledge Graphs (KG) offer easy-to-process information. An important issue to build a KG from texts is the Relation Extraction (RE) task that identifies and labels relationships between entity mentions. In this paper, to address the RE problem, we propose to combine a deep learning approach for relation detection, and a symbolic method for relation classification. It allows to have at the same time the performance of deep learning methods and the interpretability of symbolic methods. This method has been evaluated and compared with state-of-the-art methods on TACRED, a relation extraction benchmark, and has shown interesting quantitative and qualitative results.

Conceptual Navigation in Large Knowledge Graphs [40]. A growing part of Big Data is made of knowledge graphs. Major knowledge graphs such as Wikidata, DBpedia or the Google Knowledge Graph count millions of entities and billions of semantic links. A major challenge is to enable their exploration and querying by end-users. The SPARQL query language is powerful but provides no support for exploration by endusers. Question answering is user-friendly but is limited in expressivity and reliability. Navigation in concept lattices supports exploration but is limited in expressivity and scalability. In this paper, we introduce a new exploration and querying paradigm, Abstract Conceptual Navigation (ACN), that merges querying and navigation in order to reconcile expressivity, usability, and scalability. ACN is founded on Formal Concept Analysis (FCA) by defining the navigation space as a concept lattice. We then instantiate the ACN paradigm to knowledge graphs (Graph-ACN) by relying on Graph-FCA, an extension of FCA to knowledge graphs. We continue by detailing how Graph-ACN can be efficiently implemented on top of SPARQL endpoints, and how its expressivity can be increased in a modular way. Finally, we present a concrete implementation available online, Sparklis, and a few application cases on large knowledge graphs.

8.2 Interpretable Machine Learning

Participants: Hugo Ayats, Tassadit Bouadi, Peggy Cellier, Lénaïg Cornanguer, Julien Delaunay, Sébastien Ferré, Élisabeth Fromont, Luis Galárraga, Romaric Gaudel, Victor Guyomard, Christine Largouët, Véronique Masson, Laurence Rozé, Alexandre Termier.

Some of previously presented documents also contribute to this research domain: [27] [16] [18] [22] .

When Should We Use Linear Explanations? [19]. The increasing interest in transparent and fair AI systems has propelled the research in explainable AI (XAI). One of the main research lines in XAI is post-hoc explainability, the task of explaining the logic of an already deployed black-box model. This is usually achieved by learning an interpretable surrogate function that approximates the black box. Among the existing explanation paradigms, local linear explanations are one of the most popular due to their simplicity and fidelity. Despite their advantages, linear surrogates may not always be the most adapted method to produce reliable, i.e., unambiguous and faithful explanations. Hence, this paper introduces Adapted Post-hoc Explanations (APE), a novel method that characterizes the decision boundary of a black-box classifier and identifies when a linear model constitutes a reliable explanation. Besides, characterizing the black-box frontier allows us to provide complementary counterfactual explanations. Our experimental evaluation shows that APE identifies accurately the situations where linear surrogates are suitable while also providing meaningful counterfactual explanations.

XEM: An explainable-by-design ensemble method for multivariate time series classification [10]. We present XEM, an eXplainable-by-design Ensemble method for Multivariate time series classification. XEM relies on a new hybrid ensemble method that combines an explicit boosting-bagging approach to handle the bias-variance trade-off faced by machine learning models and an implicit divide-and-conquer approach to individualize classifier errors on different parts of the training data. Our evaluation shows that XEM outperforms the state-of-the-art MTS classifiers on the public UEA datasets. Furthermore, XEM provides faithful explainability by-design and manifests robust performance when faced with challenges arising from continuous data collection (different MTS length, missing data and noise).

s-LIME: Reconciling Locality and Fidelity in Linear Explanations [23], [29]. The benefit of locality is one of the major premises of LIME, one of the most prominent methods to explain black-box machine learning models. This emphasis relies on the postulate that the more locally we look at the vicinity of an instance, the simpler the black-box model becomes, and the more accurately we can mimic it with a linear surrogate. As logical as this seems, our findings suggest that, with the current design of LIME, the surrogate model may degenerate when the explanation is too local, namely, when the bandwidth parameter σ tends to zero. Based on this observation, the contribution of this paper is twofold. Firstly, we study the impact of both the bandwidth and the training vicinity on the fidelity and semantics of LIME explanations. Secondly, and based on our findings, we propose s-LIME, an extension of LIME that reconciles fidelity and locality.

VCNet: A self-explaining model for realistic counterfactual generation [25], [31]. Counterfactual explanation is a common class of methods to make local explanations of machine learning decisions. For a given instance, these methods aim to find the smallest modification of feature values that changes the predicted decision made by a machine learning model. One of the challenges of counterfactual explanation is the efficient generation of realistic counterfactuals. To address this challenge, we propose VCNet-Variational Counter Net-a model architecture that combines a predictor and a counterfactual generator that are jointly trained, for regression or classification tasks. VCNet is able to both generate predictions, and to generate counterfactual explanations without having to solve another minimisation problem. Our contribution is the generation of counterfactuals that are close to the distribution of the predicted class. This is done by learning a variational autoencoder conditionally to the output of the predictor in a joint-training fashion. We present an empirical evaluation on tabular datasets and across several interpretability metrics. The results are competitive with the state-of-the-art method.

8.3 Real World AI

8.3.1 Computer Vision and Robotics

Participants: Simon Corbille', Samuel Felton, Éliisa Fromont.

Visual Servoing in Autoencoder Latent Space [11]. Visual servoing (VS) is a common way in robotics to control a robot motion using information acquired by a camera. This approach requires to extract visual information from the image to design the control law. The resulting servo loop is built in order to minimize an error expressed in the image space. We consider a direct visual servoing (DVS) from whole images. We propose a new framework to perform VS in the latent space learned by a convolutional autoencoder. We show that this latent space avoids explicit feature extraction and tracking issues and provides a good representation, smoothing the cost function of the VS process. Besides, our experiments show that this unsupervised learning approach allows us to obtain, without labelling cost, an accurate end-positioning, often on par with the best DVS methods in terms of accuracy but with a larger convergence area.

Combination of explicit segmentation with Seq2Seq recognition for fine analysis of children handwriting [13]. We consider the task of analysing children handwriting in the context of a dictation task. The objective is to detect orthographic and phonological errors. To achieve this goal, we extend an existing handwriting analysis engine, based on an explicit segmentation of the handwritten input, originally developed for children copying exercises. We present a new approach, based on the combination of this analysis engine with a deep learning word recognition approach in order to improve both the recognition and segmentation performance. Explicit segmentation needs prior knowledge, and the deep network recognition predictions are a reliable approximation of the ground truth which can guide the analysis process. We propose to combine multiple prior knowledge strategies to further improve the analysis performance. Furthermore, we exploit the deep network approximate implicit segmentation to optimise the existing analysis process in terms of complexity.

Low-cost Multispectral Scene Analysis with Modality Distillation [35]. Despite its robust performance under various illumination conditions, multispectral scene analysis has not been widely deployed due to two strong practical limitations: 1) thermal cameras, especially high-resolution ones are much more expensive than conventional visible cameras; 2) the most commonly adopted multispectral architectures, twostream neural networks, nearly double the inference time of a regular mono-spectral model which makes them impractical in embedded environments. In this work, we aim to tackle these two limitations by proposing a novel knowledge distillation framework named Modality Distillation (MD). The proposed framework distills the knowledge from a high thermal resolution two-stream network with featurelevel fusion to a low thermal resolution one-stream network with image-level fusion. We show on different multispectral scene analysis benchmarks that our method can effectively allow the use of low-resolution thermal sensors with more compact one-stream networks.

8.3.2 Recommender Systems

Participants: Éliisa Fromont, Romaric Gaudel, Camille-Sovanneary Gauthier.

UniRank: Unimodal Bandit Algorithm for Online Ranking [24], [30]. We tackle, in the multiple-play bandit setting, the online ranking problem of assigning L items to K predefined positions on a web page in order to maximize the number of user clicks. We propose a generic algorithm, UniRank, that tackles state-of-the-art click models. The regret bound of this algorithm is a direct consequence of the unimodality-like property of the bandit setting with respect to a graph where nodes are ordered sets of indistinguishable items. The main contribution of UniRank is its $O(L/\Delta \log T)$ regret for T consecutive assignments, where Δ relates to the reward-gap between two items. This regret bound is based on the usually implicit condition that two items may not have the same attractiveness. Experiments against state-of-the-art learning algorithms specialized or not for different click models, show that our method has better regret performance than other generic algorithms on real life and synthetic datasets.

8.3.3 Forecasting

Participants: Abderaouf Nassim Amalou, Éliisa Fromont.

CATREEN : Context-Aware Code Timing Estimation with Stacked Recurrent Networks [15], [43]. Automatic prediction of the execution time of programs for a given architecture is crucial, both for performance analysis in general and for compiler designers in particular. In this paper, we present CATREEN, a recurrent neural network able to predict the steady-state execution time of each basic block in a program. Contrarily to other models, CATREEN can take into account the execution context formed by the previously executed basic blocks which allows accounting for the processor micro-architecture without explicit modeling of micro-architectural elements (caches, pipelines, branch predictors, etc.). The evaluations conducted with synthetic programs and real ones (programs from Mibench and Polybench) show that CATREEN can provide accurate prediction for execution time with 11.4% and 16.5% error on average, respectively and that we got an improvement of 18% and 27.6% respectively when comparing our tool estimations to the state-of-the-art LSTM-based model.

8.3.4 Logistics

Participants: Éliisa Fromont, Gregory Martin, Laurence Rozé, Alexandre Termier.

Optimisation du Positionnement de Voitures en Autopartage basée sur la Prédiction de leur Utilité [32]. The success of a free-floating car-sharing service depends on a good allocation of the vehicles across the city, i.e., where and when they are needed by citizens. This requires predicting the demand across the geographical regions and across time, which is challenging due to the sparsity and variability of the data. Furthermore, the purpose of these predictions is to help compute the best possible car relocation for the next day, hence the need to model both the prediction task and the optimization task in a compatible way. As the allocation optimization involves reasoning about the number of cars to assign to geographical regions, we propose to convert the prediction problem into predicting the expected utilization of a car when added to a region. We discuss the challenges in modeling both the machine learning and the relocation problem, and we propose a mixed-integer linear programming method that solves the relocation problem while taking into account the model predictions and relocation distances. We experiment with two datasets from citywide car sharing companies and show how our method can increase the allocation strategies and hence profitability of the services.

8.3.5 Privacy

Participants: Élisabeth Fromont, Antonin Voyez.

Membership Inference Attacks on Aggregated Time Series with Linear Programming [34]. Aggregating data is a widely used technique to protect privacy. Membership inference attacks on aggregated data aim to infer whether a specific target belongs to a given aggregate. We propose to study how aggregated time series data can be susceptible to simple membership inference privacy attacks in the presence of adversarial background knowledge. We design a linear programming attack that strongly benefits from the number of data points published in the series and show on multiple public datasets how vulnerable the published data can be if the size of the aggregated data is not carefully balanced with the published time series length. We perform an extensive experimental evaluation of the attack on multiple publicly available datasets. We show the vulnerability of aggregates made of thousands of time series when the aggregate length is not carefully balanced with the published length of the time series.

Unique in the Smart Grid -The Privacy Cost of Fine-Grained Electrical Consumption Data [45]. The collection of electrical consumption time series through smart meters grows with ambitious nationwide smart grid programs. This data is both highly sensitive and highly valuable: strong laws about personal data protect it while laws about open data aim at making it public after a privacy-preserving data publishing process. In this work, we study the uniqueness of large scale real-life fine-grained electrical consumption time-series and show its link to privacy threats. Our results show a worryingly high uniqueness rate in such datasets. In particular, we show that knowing 5 consecutive electric measures allows to re-identify on average more than 90% of households in our 2.5M half-hourly electric time series dataset. Moreover, uniqueness remains high even when data is severely degraded. For example, when data is rounded to the nearest 100 watts, knowing 7 consecutive electric measures allows to re-identify on average more than 40% of the households (same dataset). We also study the relationship between uniqueness and entropy, uniqueness and electric consumption, and electric consumption and temperatures, showing their strong correlation.

8.3.6 Medicine

Participants: Élisabeth Fromont, Elodie Germani.

On the benefits of self-taught learning for brain decoding [44]. We study the benefits of using a large public neuroimaging database composed of fMRI statistic maps, in a self-taught learning framework, for improving brain decoding on new tasks. First, we leverage the NeuroVault database to train, on a

selection of relevant statistic maps, a convolutional autoencoder to reconstruct these maps. Then, we use this trained encoder to initialize a supervised convolutional neural network to classify tasks or cognitive processes of unseen statistic maps from large collections of the NeuroVault database. We show that such a self-taught learning process always improves the performance of the classifiers but the magnitude of the benefits strongly depends on the number of data available both for pre-training and finetuning the models and on the complexity of the targeted downstream task.

8.3.7 Agriculture

Participants: Tassadit Bouadi, Olivier Gauriau, Christine Largouët, Véronique Masson, Alexandre Termier.

Agriculture and Digital Technology: Getting the most out of digital technology to contribute to the transition to sustainable agriculture and food systems [36], [37], [39], [38], [41]. This white paper is an initiative of the executive boards of INRAE and Inria, who brought us together, tasked us and gave us carte blanche in coordinating and writing this work on digital technology in agriculture from a research perspective. Our reflections have been carried out in the context of the principal observed and foreseeable dynamics of agriculture in countries around the world that aim to support the development of more sustainable agricultural and food systems. In constructing the analysis and proposals presented in this publication, based on an inventory of the challenges, opportunities and risks associated with digital technology in agriculture, we have not aimed to simply focus on the consensus of opinion, but to express the diversity of the groups involved, in particular in terms of prioritising the challenges or scope of the dangers for agriculture and food systems of the future. This white paper is a collective work that aims to address the question of digital technology in agriculture as a balanced whole and its chapters are sequenced and interconnected in a way that allows all aspects to be covered. Each section should therefore be considered in perspective of the others, and using them in isolation without referring to the whole could present an unbalanced view.

The french INRIA- INRAE White Book “Agriculture and digital technologies” - Coupling digitalization and agroecology to design more sustainable and resilient food systems [28]. Digital technologies are spreading in agriculture as they do in other economic sectors. Facing this wave, two major french research institutes dedicated to agriculture (INRAE) and digital sciences (INRIA) adress the following question « how can digital technologies be designed to accelerate the transition towards more sustainable and resilient food systems, including agroecology, climate-change resilient agriculture and adaptation to food transitions ? » in a White Book, published in 2022 that is presented in the webinar. The approach has been to review the opportunities offered by the state of the art of digital technologies and the risks linked to these technologies, and to confront them in order to design research avenues that enable us to take advantage of the opportunities while mitigating risks. Opportunities have been found in 3 directions : better agricultural production, better inclusion in value chains and networks, and better knowledge sharing. Risks are economical (cost, power control), social (digital divide, deskilling), technical (cybersecurity, complexity), ecological (ICT footprint...) and also linked to sovereignty regarding data. Eventually, research avenues have been built in four areas : (1) digital technologies for creating and sharing data and knowledge, (2) digital technologies for helping farmers in farm management, (3) digital technologies for accompanying the collective management of territories and (4) digital technologies for better inclusion in the value chains. These areas of research are crossed by four transversal challenges, that will guide research : (1) developping holistic approaches, (2) searching resilience and not optimum, (3) looking for frugal solutions (green IT) and (4) ensuring confidence of the users (security, transparency). Last, four final messages will be delivered to pave the way to future research for a resonsible digital agriculture, in France and abroad.

Precision feeding of lactating sows: implementation and evaluation of a decision support system in farm conditions [12]. Precision feeding (PF) aims to provide the right amount of nutrients at the right time for each animal. Lactating sows generally receive the same diet, which either results in insufficient

supply and body reserve mobilization, or excessive supply and high nutrient excretion. With the help of online measuring devices, computational methods, and smart feeders, we introduced the first PF decision support system (DSS) for lactating sows. Precision (PRE) and conventional (STD) feeding strategies were compared in commercial conditions. Every day each PRE sow received a tailored ration that had been computed by the DSS. This ration was obtained by blending a diet with a high AA and mineral content (13.00 g/kg SID Lys, 4.50 g/kg digestible P) and a diet low in AAs and minerals (6.50 g/kg SID Lys, 2.90 g/kg digestible P). All STD sows received a conventional diet (10.08 g/kg SID Lys, 3.78 g/kg digestible P). Before the trial, the DSS was fitted to farm performance for the prediction of piglet average daily gain (PADG) and sow daily feed intake (DFI), with data from 1,691 and 3,712 lactations, respectively. Sow and litter performance were analyzed for the effect of feeding strategy with ANOVA, with results considered statistically significant when ± 0.05 . The experiment involved 239 PRE and 240 STD sows. DFI was similarly high in both treatments (PRE: 6.59, STD: 6.45 kg/d; $P=0.11$). Litter growth was high (PRE: 2.96, STD: 3.06 kg/d), although it decreased slightly by about 3% in PRE compared to STD treatments (± 0.05). Sow body weight loss was low, although it was slightly higher in PRE sows (7.7 versus 2.1 kg, ± 0.001), which might be due to insufficient AA supply in some sows. Weaning to estrus interval (5.6 d) did not differ. In PRE sows SID Lys intake (PRE: 7.7, STD: 10.0 g/kg; ± 0.001) and digestible P intake (PRE: 3.2, STD: 3.8 g/kg; ± 0.001) declined by 23% and 14%, respectively, and feed cost decreased by 12%. For PRE sows, excretion of N and P decreased by 28% and 42%, respectively. According to these results, PF appears to be a very promising strategy for lactating sows.

XPM: An explainable-by-design pattern-based estrus detection approach to improve resource use in dairy farms [26]. A powerful automatic detection of estrus, the only period when the cow is susceptible to pregnancy, is a key driver to help farmers with reproduction management and subsequently to improve milk production resource use in dairy farms. Automatic solutions to detect both types of estrus (behavioral and silent estrus) based on the combination of affordable phenotyping data (activity, body temperature) exist, but they do not provide faithful explanations to support their alerts and in ways that farmers can understand based on the behaviors they could observe in animals. In this paper, we first propose XPM, a novel pattern-based classifier to detect both types of estrus with real-world affordable sensor data (activity, body temperature) which supports its predictions with perfectly faithful explanations. Then, we show that our approach performs better than a commercial reference in estrus detection, driven by the detection of silent estrus. Finally, we present the explainability of our solution which stems from the communication to the farmers the presence and/or absence of a limited number of patterns determinant of estrus detection, therefore reducing solution mistrust and supporting farmers' decision-making.

Prédiction des épidémies de cercosporiose de la betterave par des approches d'apprentissage automatique. [33]. Cercosporiosis is a major beet disease with up to 30% yield loss and a decrease in sugar content of 1 to 2 points in case of uncontrolled epidemics. The objective of our work was to build predictive decision support tools by exploiting the numerous data on cercosporiosis in beet from the epidemiological monitoring networks for the Bulletin de Santé du Végétal and collected since 2009 in the northern part of France, using a machine learning approach. First, we tried to predict the date of appearance of the first symptoms in order to have information on the beginning of the epidemic season beforehand (mid-June). Secondly, we wanted to predict the epidemic dynamics during the season by predicting the evolution of the disease over the coming week (D+7) to better manage the control of cercosporiosis.

9 Bilateral contracts and grants with industry

9.1 Bilateral contracts with industry

- ORANGE - Univ. Rennes I

Participants: Tassadit Bouadi, Alexandre Termier, Victor Guyomard.

Contract amount: 30k€ + Phd Salary

Context. This project is a collaboration with Orange Labs Lannion about interpretable machine learning. The Orange company aims to develop the use of machine learning algorithms to enhance the services they propose to their customers (for instance, credit acceptance or attribution prediction). It ensues the development of *generic approaches for providing interpretable decisions* to customers or client managers.

Objective. The GDPR, implemented by the EU in 2018, stipulates the right for explanations for EU citizens in regards to decisions made from personal data. In a society where many of those decisions are computer-assisted via machine learning algorithm, interpretable ML is crucial. A promising way to convey explanations for the outcomes of ML models are *counterfactual explanations*. The focus of the PhD thesis financed by this project is the generation of usable and actionable counterfactual explanations for ML classifiers, which are intensively used by Orange within their services.

Additional remarks. This contract finances the PhD of Victor GUYOMARD by Orange.

- **Louis Vuitton - Univ. Rennes I**

Participants: Romaric Gaudel, Elisa Fromont, Camille Sovanneary Gauthier.

Participants: R. Gaudel, E. Fromont, C. Sovanneary Gauthier

Contract amount: 60k€ (shared with CREST)

Context. Louis Vuitton is a French high-end luxury fashion house with a large catalog of products available for purchase on their site web via keyword research. Hence, pertinence of search results as well as useful recommendations are of paramount importance for their business.

Objective. To tackle the particular recommendation use cases encountered at Louis Vuitton, we focus on the *online learning to rank* problem: we identify the right click behavioral models for clients and we develop new bandit algorithms to efficiently infer the parameters of such click behavioral models.

Additional remarks. This contract finances the PhD of Camille-Sovanneary Gauthier.

- **PSA - Inria**

Participants: Elisa Fromont, Alexandre Termier, Laurence Rozé, Grégory Martin.

Contract amount: 75k€

Context. Peugeot-Citroën (PSA) group aims at improving the management of its car sharing service. To optimize its fleet and the availability of the cars throughout the city, PSA needs to analyze the trajectory of its cars.

Objective. The aim of the internship is (1) to survey the existing methods to tackle the aforementioned need faced by PSA and (2) to also investigate how the techniques developed in LACODAM (e.g., emerging pattern mining) could be serve this purpose. A framework, consisting of three main modules, has been developed. We describe the modules in the following.

- A town modelisation module with clustering. Similar towns are clustered in order to reuse information from one town in other towns.
- A travel prediction module with basic statistics.
- A reallocation strategy module (choices on how to relocate cars so that the most requested areas are always served). The aim of this module is to be able to test different strategies.

Additional remarks. This is the doctoral contract for the PhD of Gregory Martin (Thèse CIFRE).

10 Partnerships and cooperations

10.1 International research visitors

10.1.1 Visits of international scientists

Other international visits to the team

Gonzalo Méndez

Status: Researcher

Institution of origin: ESPOL³

Country: Ecuador

Dates: 31/03 - 30/04 and 09/11 - 13/12

Context of the visit: This research visit happened in the context of an ongoing collaboration between the Luis Galárraga and Gonzalo Méndez. This collaboration started in 2019 with an internship co-supervision on the topic of AI-based academic advising. The fruit of this joint work is the design of the iCoRA system [46], a visual tool for AI-based course recommendation for students at ESPOL. The deployment of iCoRA has set the ground to study the human aspects of conveying AI-based predictions (and explanations) to human users. In a new study we investigated the effects of different visual representations for predictions on the decisions and behavior of the students. The results of this study have been already published [48] at CHI 2021 and Gonzalo's first visit aimed to continue this line of research, by (a) exploiting eye tracking technologies to gather more precise data, and (b) to study the problem from the perspective of the academic advisors – who will also use the system. Moreover, his first visit allowed us to build a new collaboration axis around the visualization of the predictions and explanations of rule-based regression models applied to predicting the incidence of plant diseases (doctoral work of Olivier Gauriau). These new projects have been addressed more deeply during his second visit. Fruits of this new visit are a submitted paper to CHI 2023 – a follow-up of our work of 2021 –, and a visualization prototype that will be validated by a group of agronomists working with disease monitoring for cultures of beetroot and vine.

Mobility program/type of mobility: Collaboration LACODAM/ESPOL (first visit), Bourse de Séjour Scientifique de Haut Niveau (second visit)

Assie Brou Ida

Status: Associate Professor

Institution of origin: University Félix Houphouët-Boigny, Abidjan

Country: Ivory Cost

Dates: 30/05/2022 - 17/06/2022

Context of the visit: In the context of the cooperation between University of Rennes 1 and University Félix Houphouët-Boigny, a double diploma MIAGE Master is being implemented between the two institutions. Dr Assie will teach in this master: the goal of her visit was to increase her expertise in the data science area in order to teach modules in this future master. We could also exchange on her research work, which concerned the analysis of interview data in collaboration with psychologists.

Mobility program/type of mobility: Cooperation between University of Rennes 1 and University Félix Houphouët-Boigny

³Escuela Superior Politécnica del Litoral

10.1.2 Visits to international teams

Research stays abroad

Julien Delaunay

Visited institution: Aalborg University

Country: Denmark

Dates: 15/08 - 22/12

Context of the visit: Julien visited the Human-Centered Computing group of the CS Department of Aalborg University as a guest PhD student under the supervision of **Dr. Niels van Berkel**. This inter-disciplinary collaboration lies at the cross-roads of eXplainable AI and Data Visualization and is particularly concerned with understanding the impact of the explanation paradigm (feature-attribution, rules, and counterfactual explanations) and the visual representation (charts vs. text) on the comprehensibility and perceived trustworthiness of explanations for AI agents. To this end Julien has designed and conducted – under the joint supervision of his doctoral advisors and Niels – an experimental protocol based on a user study that will shed light on which ways to convey explanations for AI models are the most effective from an end-user perspective. This project, as well as Julien’s doctoral work, is part of the FabLe ANR ported by Luis Galárraga.

Mobility program/type of mobility: MATHStics Mobility Scholarship

10.2 European initiatives

10.2.1 H2020 projects

- **TAILOR: Foundations of Trustworthy AI – Integrating Reasoning, Learning and Optimization**

Participants: Éliisa Fromont, Alexandre Termier, Luis Galárraga.

TAILOR is an EU project with the aim build the capacity to provide the scientific foundations for Trustworthy AI in Europe. TAILOR develops a network of research excellence centres, leveraging and combining learning, optimisation, and reasoning. These systems are meant to provide descriptive, predictive, and prescriptive systems integrating data-driven and knowledge-based approaches.

Dates: 09/2020 - 08/2024

10.3 National initiatives

- **HyAIAI: Hybrid Approaches for Interpretable AI**

Participants: Elisa Fromont (leader), Alexandre Termier, Luis Galárraga, Neetu Kushwaha, Ezanin Bile.

The Inria Project Lab HyAIAI is a consortium of Inria teams (Sequel, Magnet, Tau, Orpailleur, Multispeech, and LACODAM) that work together towards the development of novel methods for machine learning, that combine numerical and symbolic approaches. The goal is to develop new machine learning algorithms such that (i) they are as efficient as current best approaches, (ii) they can be guided by means of human-understandable constraints, and (iii) their decisions can be better understood.

- **#DigitAg: Digital Agriculture**

Participants: Alexandre Termier, Véronique Masson, Christine Largouët, Luis Galárraga, Olivier Gauriau.

#DigitAg is a “Convergence Institute” dedicated to the increasing importance of digital techniques in agriculture. Its goal is twofold: First, making innovative research on the use of digital techniques in agriculture in order to improve competitiveness, preserving the environment, and offer correct living conditions to farmers. Second, preparing future farmers and agricultural policy makers to successfully exploit such technologies. While #DigitAg is based on Montpellier, Rennes is a satellite of the institute focused on cattle farming.

LACODAM is involved in the “data mining” challenge of the institute, which A. Termier co-leads. He is also the representative of Inria in the steering committee of the institute. The interest for the team is to design novel methods to analyze and represent agricultural data, which are challenging because they are both heterogeneous and multi-scale (both spatial and temporal).

- **PEPR WAIT 4**

Participants: Alexandre Termier, Peggy Cellier, Lucie Lepetit, Christine Largouët, Véronique Masson, Louis Bonneau De Beaufort.

The WAIT 4 project is a part of the “Agroecology and numeric” PEPR. The goal of this project is to provide the scientific basis for significant improvements in the well-being of farm animals. Up to now, animal well-being is evaluated with indicators of the means deployed (e.g. available space, method to control building temperature, time spent outside...). The goal of WAIT4 is to provide tools required in order to move to *results* indicators: can some guarantees be given on the well being of animals? Can this well (or unwell) being be correlated to management actions from the farmer, or to their general living conditions?

This requires a much finer understanding of animal mental as well as physiological state. The project is lead by Inrae (Florence Gondret), which brings animal science specialists, ranging from biologists to ethologists. CEA provides expertise on blood sensors, to measure molecules linked to stress. And Inria as well as Insa Lyon provide computer science expertise for tools to analyse the data. More precisely, the Lacodam team will deal first with analyzing time series of numerical sensor data (e.g. temperature, activity), and second with categorical sequences of events produced by annotation tools from the analysis of videos. Both will help to better model animal behavior, and determine what are “normal” behaviors, and what are anomalous behaviors that may be linked to bad conditions for the animals.

- **Bourse IUF - Elisa FROMONT**

This project supports the work of Elisa Fromont both with a reduction of teaching load, and some research money (15k). Elisa is currently working on designing effective data mining and machine learning algorithms for real-life data (which are scarce, heterogenous, multimodal, imbalanced, temporal, ...). For the next few years, Elisa would like to focus on the interpretability of the results obtained by these algorithms. In pattern mining, her goal is to design algorithms which can directly mine a small number of relevant patterns. In the case of black box machine learning models (e.g. deep neural nets), Elisa would like to design methods to help the end user understand the decisions taken by the model.

- **Scikit-mine (F-WIN project of PNR-IA)**

Participants: Peggy Cellier, Alexandre Termier.

Scikit-mine (SKM for short) is a Python library of pattern mining algorithms, designed to be compatible with the well-known scikit-learn library. It allows practitioners to use state-of-the-art pattern mining algorithm with a library that has the same usage interface as scikit-learn, and that exploits the same data types. SKM is currently developed by CNRS AI engineers in the context of the F-WIN project of the PNR-IA program of CNRS, which general goal is to improve the development of AI software in research teams of CNRS labs. The goal of this project is to make SKM robust enough for a public release in 2023.

10.3.1 ANR

- **FAbLe: Framework for Automatic Interpretability in Machine Learning**

Participants: L. Galárraga (holder), C. Largouët

Participants: Luis Galárraga (holder), Christine Largouët, Julien Delaunay.

How can we fully automatically choose the best explanation for a given use case in classification? Answering this question is the raison d'être of the JCJC ANR project FAbLe. By "best explanation" we mean the explanation that yields the best trade-off between interpretability and fidelity among a universe of possible explanations. While fidelity is well-defined as the accuracy of the explanation w.r.t the answers of the black-box, interpretability is a subjective concept that has not been formalized yet. Hence, in order to answer our prime question we first need to answer the question: "How can we formalize and quantify interpretability across models?". Much like research in automatic machine learning has delegated the task of accurate model selection to computers [47], FAbLe aims at fully delegating the selection of interpretable explanations to computers. Our goal is to produce a suite of algorithms that will compute suitable explanations for ML algorithms based on our insights of what is interpretable. The algorithms will choose the best explanation method based on the data, the use case, and the user's background. We will implement our algorithms so that they are fully compatible with the body of available software for data science (e.g., Scikit-learn).

- **SmartFCA: A Smart Tool for Analyzing Complex Data with Formal Concept Analysis**

Participants: Sébastien Ferré, Peggy Cellier.

Period: 01/01/2022 – 31/12/2025

Budget: 143k€ (Univ Rennes)

Formal Concept Analysis (FCA) is a mathematical framework based on lattice theory and aimed at data analysis and classification. FCA, which is closely related to pattern mining in knowledge discovery (KD), can be used for data mining purposes in many application domains, e.g. life sciences and linked data. Moreover, FCA is human-centered and provides means for visualization and interaction with data and patterns. Actually it is now possible to deal with complex data such as intervals, sequences, trajectories, trees, and graphs. Research in FCA is dynamic, but there is still room for extensions of the original formalism. Many theoretical and practical challenges remain. Actually there does not exist any consensual platform offering the necessary components for analyzing real-life data. This is precisely the objective of the SmartFCA project to develop the theory and practice of FCA and its extensions, to make the related components inter-operable, and

to implement a usable and consensual platform offering the necessary services and workflows for KD.

In particular, for satisfying in the best way the needs of experts in many application domains, SmartFCa will offer a “Knowledge as a Service” (KaaS) component for making domain knowledge operable and reusable on demand.

- **MeKaNo: Search the Web with Things**

Participants: Sébastien Ferré, Peggy Cellier, Luis Galárraga, Julie Boudebs.

Period: 01/10/2022 – 29/09/2026

Budget: 143k€ (Univ Rennes)

In MeKaNo, we aim to search the web with things, in order to get more accurate results over a wide diversity of sources. Traditional web search engines search the web with strings. However, keyword search often returns many irrelevant documents, pushing users to refine their keyword list following a trial-and-error process. To overcome such limitations, major companies allowed searching for things, not strings. Asking for the age of “James Cameron” to your vocal assistant, it locates in a Knowledge Graph (KG) a Person matching “James Cameron” where a property “age” is set to 66 years, i.e. the Thing “James Cameron”. If searching for Things is a tremendous progress and delivers exact answers, the search is done over a Knowledge Graph and not on the Web. Consequently, there may exist many answers on the web that are not part of the knowledge graph.

To summarize, searching with strings over the web offers diversity at the expense of noise. Searching for Things delivers exact answers, but we lose diversity. In MeKaNo, we aim at searching the web with Things to get diversity and avoid noisy results. To search the web with Things, we face three main scientific challenges:

1. Users are used to search with keywords. Transforming a keyword query into a mixed query that first searches over a KG then into the web is difficult, especially, for complex queries.
2. As with traditional web searches, users expect to obtain ranked results in a snap. Combining KG search and Web search while preserving performances is highly challenging and requires a new kind of search engine.
3. Improving the connection between the web of microdata and Knowledge Graphs requires entity matching at large scale for microdata entities and KG entities.

10.4 Regional initiatives

- **MiKroloG: The Microdata Knowledge Graph (Labex CominLabs)**

Participants: Sébastien Ferré, Peggy Cellier, Julie Boudebs.

Period: 01/09/2021 - 31/12/2024

Budget: 56k€ (Univ Rennes)

Today, a few keywords are enough to find a relevant document among the billions of documents scattered on the web. If search engines allow us to apprehend the diversity of the web, they also confront us with its noise and the search for relevant information can be more difficult than expected. A few keywords are also the necessary information to find concepts in knowledge graphs. *Who is the author of Twenty Thousand Leagues Under the Sea? How far is the earth from the moon?*

Voice assistants simply allow us to access controlled and verified knowledge, but this is no longer the web with all its diversity.

In MiKroloG, we are looking to establish a closer link between knowledge graphs and the web. The idea is to be able to search for documents on the web but starting from concepts identified in knowledge graphs. In this way, we hope to reconcile the high precision of knowledge graphs with the diversity of the web. To achieve this, MiKroloG focuses on three main scientific challenges: (i) resolving the entities between the web of micro-data and knowledge graphs, (ii) evaluating queries on large knowledge graphs and ranking the results, (iii) expressing complex queries to end-users.

11 Dissemination

11.1 Promoting scientific activities

Participants: the whole team.

11.1.1 Scientific events: organisation

- Elisa Fromont was General chair of the Symposium on Data Analysis ([IDA 2022](#)). All LACODAM (including Gaëlle Tworkowski, the project assistant) was involved in the organization of this event.
- Tassadit Bouadi and Luis Galárraga were co-organizers of the workshop on Advances in Interpretable Machine Learning and Artificial Intelligence ([AIMLAI](#)) co-located with the International Conference on Knowledge Management ([CIKM 2022](#)).

11.1.2 Scientific events: selection

Chair of conference program committees

- Tassadit Bouadi was Program Chair of the Symposium on Data Analysis ([IDA 2022](#))
- Elisa Fromont was co-program chair of the “Conférence sur l’Apprentissage automatique” ([CAp 2022](#)) in Vannes.
- Romaric Gaudel is co-program chair of the “Conférence sur l’Apprentissage automatique” ([CAp 2023](#)) in Strasbourg.
- Peggy Cellier was co-Journal track chair of the European Conference in Machine Learning and Knowledge Discovery [ECML PKDD 2022](#) in Grenoble

Member of the conference program committees

- Elisa Fromont was in the program committee of KDD 2022 ("ACM SIGKDD Conference on Knowledge Discovery and Data Mining"); ECMLPKDD 2022 and IDA 2022 ("European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases" and "Intelligent Data Analysis") as senior PC; IJCAI 2022 ("International Joint Conference in AI"); WACV 2022 (IEEE's and the PAMI-TC's "Winter Conference on Applications of Computer Vision").
- Peggy Cellier was in the program committee of ECMLPKDD 2022 ("European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases") as Area chair; EGC 2022 ("Extraction et Gestion de la Connaissances") as senior PC; ICCS 2022 ("International Conferences on Conceptual Structures"); ICFCA 2022 ("International Conference on Formal Concept Analysis"); CLA 2022 ("Concept Lattices and Applications"); ETAFCA 2022 ("Workshop on Existing Tools and Applications for Formal Concept Analysis"); FCA4AI 2022 ("Workshop on FCA"); TALN 2022 ("Conférence sur le Traitement Automatique des Langues Naturelles").

- Luis Galárraga was in the program committee of IJCAI 2022 (International Joint Conference in AI); WSDM 2022 (ACM International Conference on Web Search and Data Mining); Posters/Demo Session at EKAW 2022 (International Conference on Knowledge Engineering and Knowledge Management); ISWC 2022 (International Semantic Web Conference); Posters/Demo Session at ESWC 2022 (Extended Semantic Web Conference); Wikidata 2022 (Wikidata Workshop).
- Romaric Gaudel was in the program committee of ICML 2022 (International Conference on Machine Learning); IDA 2022 (Symposium on Data Analysis); CAp 2022 ("Conférence sur l'Apprentissage automatique").
- Alexandre Termier was in the program committee of KDD 2022 ("ACM SIGKDD Conference on Knowledge Discovery and Data Mining"); SDM 2022 ("SIAM Data Mining Conference"); IDA 2022 ("Intelligent Data Analysis Conference")
- Sébastien Ferré was in the program committee of WWW 2022 ("The Web Conference"); ESWC 2022 ("Extended Semantic Web Conference"); IDA 2022 ("Intelligent Data Analysis"); CLA 2022 ("Concept Lattices and Applications").
- Tassadit Bouadi was in the program committee of IDA 2022 ("Intelligent Data Analysis Conference") and ECMLPKDD 2022 ("European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases")

11.1.3 Journal

Member of the editorial boards

- Luis Galárraga (in collaboration with Miguel Couceiro) is a guest editor for the special issue "Fair and Explainable Models" at the EURO Journal of Decision Processes.
- Elisa Fromont is Co-Specialty Chief Editor of Frontiers in Artificial Intelligence specialty Machine Learning and Artificial Intelligence.
- Alexandre Termier is Member of the Editorial Board of the Data Mining and Knowledge Discovery journal (DMKD).
- Elisa Fromont is a guest editor of the Mathematics journal for a special issue on Time Series Analysis.
- Peggy Cellier is a Member of the Editorial Board of ICFCA ("International Conference on Formal Concept Analysis").

Reviewer - reviewing activities

- Luis Galárraga served as a reviewer for the Journal on Web Semantics (JoWS); the Semantic Web Journal (SWJ); the Knowledge and Information Systems (KIS); The Journal of Data Mining and Knowledge Discovery (DAMI); The Machine Learning Journal (MACH) The Web Conference 2023; KDD 2022 (ACM SIGKDD Conference on Knowledge Discovery and Data Mining).
- Romaric Gaudel served as a reviewer for Transactions on Machine Learning Research (TMLR).
- Christine Largouët served as a reviewer for BioSystems and Computational Biology.
- Alexandre Termier served as a reviewer for the Data Mining journal (DMKD)
- Sébastien Ferré served as a reviewer for Artificial Intelligence Review (AIRE).

11.1.4 Invited talks

Elisa Fromont did the following invited talks:

- 16/09/2022: Invited Talk "Computing How-Provenance for SPARQL Queries via Query Rewriting" for the internal seminar of the **Links** team at Inria Lille.
- 15/12/2022: Keynote (in French) at Technoférence, pôle de compétitivité I&R, "Panel des domaines de l'IA et challenges actuels de l'apprentissage automatique – frugalité, adaptabilité, confiance", Virtual.
- 28/11/2022: 2h Training session (in French) at Rennes Métropole "Introduction to AI and Machine Learning", Rennes.
- 26/11/2022: Invited talk (in French) for high school (female) students "Intelligence Artificielle de quoi parle-t-on ?", RJMI (Rendez-vous des Jeunes Mathématiciennes et Informaticiennes).
- 17/11/2022: Keynote at CIIA2022 (3era Conferencia Interdisciplinaria en Inteligencia Artificial) on "Explainable Time Series Classification", Mexico (virtual).
- 14/10/2022: Keynote (in French) at SIFED (Symposium International Francophone sur l'Ecrit et le Document) on "Explainable Time Series Classification", Rennes.
- 12/07/2022: Invited speaker (in French) at HELP workshop in GDR Madics on "Explainable Time Series Classification", Virtual.
- 11/07/2022: Presentation for the EIT Digital summer school "Smart Cities" as "Business Case Provider" on "Privacy threats when publishing public power consumption time series" in Rennes.
- 19/05/2022: Invited talk for an internal seminar at IRISA ("LinkMedia Speaks Science") on "Multi-spectral Object Detection", Rennes.
- 25/03 & 16/05 2022: Invited talk (in French) for high school students "Intelligence Artificielle de quoi parle-t-on ?", Rennes.
- 26/01/2022: Panel speaker at EGC'22 (Conférence francophone sur l'Extraction et la Gestion des Connaissances) about "Are we D&I ?", Virtual, FR. A journal article was published after this panel.

Alexandre Termier did the following invited talks:

- 25/04/2022: Keynote (in French) at the Ecole Internationale de Recherche (EIR) Agreenium about "Transition numérique, Intelligence Artificielle : de quoi parle-t-on ? quels enjeux pour la recherche et les secteurs agricoles et alimentaires ?", Nantes, FR.
- 23/03/2022: Keynote (in French) at SmartAgri days about "L'IA en agriculture: moissonner les données pour quels fruits?", Pommerit, FR.
- 26/01/2022: Panel speaker at EGC'22 (Conférence francophone sur l'Extraction et la Gestion des Connaissances) about "Are we D&I ?", Virtual, FR. A journal article was published after this panel.

Romarc Gaudel did the following invited talk:

- 28/06/2022: Tutorial (in French) "Deux catégories de systèmes de recommandation : le filtrage collaboratif et les bandits manchots" at PFIA 2022 (Plate-Forme Intelligence Artificielle), Saint-Étienne, France.

Sébastien Ferré did the following invited talk:

- 28/06/2022: Invited speaker at the "Journée DECADE" colocated with PFIA 2022 in St Etienne, titled "Instance-based Reasoning in Knowledge Graphs with Concepts of Neighbors".

11.1.5 Leadership within the scientific community

- Elisa Fromont is member of the steering committee of the “Conférence sur l’Apprentissage automatique” (CAp since 2020).
- Peggy Cellier is member of the steering committee of the European Conference in Machine Learning and Knowledge Discovery (ECML PKDD) since 2022.

11.1.6 Scientific expertise

- Elisa Fromont is a member of the IUF board, of the CSV ("Comité de Sélection et de Validation") of Images & Réseaux, of the scientific council of the GDR IA, of the Machine Learning College at AFIA, and of the “Société Savante Francophone d’Apprentissage Machine” (SSFAM).
- Alexandre Termier was a member of core team of 5 authors that prepared Inrae-Inria white book on digital agriculture. In the team, Tassadit Bouadi, Véronique Masson and Christine Largouët all contributed to the writing of this book.
- Sébastien Ferré was a member of the scientific committee of ABES, the Agency of Libraries in Higher Education, as an expert in Semantic Web technologies.
- Romaric Gaudel was reviewer for the ANR.
- Tassadit Bouadi is member of the research group of the Le Programme "Intelligence environnementale" Commun (PIEC), jointly coordinated by the MSHB (Maison des Sciences de l’Homme en Bretagne) and the OSUR (Observatoire de Rennes)
- Christine Largouët is member of the Scientific and Technological Programm Council (CSTP) of the PEPR "Agroecology and ICT" (2022-2029)
- Sébastien Ferré was an external evaluator for the Research Fund of Quebec in Nature and Technologies (FRQNT).

Hiring committees

- Elisa Fromont was a member of a hiring committee in Saint-Etienne PU27&61-4369, Rennes PU27-4078, Rennes MCF27-1575, Nantes PU27-0128, TelecomParis MdC27&61, CentraleSupélec 27CPI, Paris13 Repyramidage, Rennes Repyramidage.
- Alexandre Termier was member of a professor recruitment committee in Lannion (ENSSAT, Université de Rennes). He also reviewed “dispense de qualification” documents for MdC recruitment in Grenoble Alpes University.
- Peggy Cellier was member of two associate professor recruitment committees: in Annecy (IUT d’Annecy) and in Rennes (INSA Rennes).
- Sébastien Ferré was the president of the recruitment committee for an associate professor position at ISTIC, University of Rennes 1.

11.1.7 Research administration

- Since September 2018, Peggy Cellier is in charge of the Irisa Ph.D. students at IRISA, i.e. she is involved in the "commission du personnel" and organizes the selection of Ph.D. students for ministerial grants (contrats doctoraux). She is also an elected member of the “Conseil de Composante IRISA/INSA” at INSA. Peggy Cellier is "secrétaire" of "Revue de Traitement automatique des langues" since 2019.

- Since 2017, Elisa Fromont is elected at IRISA lab council (she is a member of the gender equality group and responsible for the anti-harassment group). She is elected, since 2020, at the scientific council of the University (UR1) and is a member of the HDR committee for the University. She is the head of the D7 scientific department at IRISA since Sept 2021 (and part of the direction scientific board).
- Since 2015, Christine Largouët is member of the scientific department council (COREGE) at Institut Agro Rennes Angers.

11.2 Teaching - Supervision - Juries

11.2.1 Teaching

Apart from Luis Galárraga (research scientist) and Gaëlle Tworowsky (administrative assistant), each permanent member of the project-team LACODAM is also faculty members and is actively involved in computer science teaching programs in ISTIC, IUT of Lannion, INSA, or Agrocampus-Ouest. Besides these usual teachings LACODAM is responsible of some teaching tracks and of some courses.

Teaching tracks responsibility

- Veronique Masson is the head of the L3 studies in Computer Science at University of Rennes 1
- Since September 2021, Alexandre Termier is co-head of Master 2 SIF (Science Informatique - research master in Computer Science) at University of Rennes 1, with Bertrand Couasnon (INSA Rennes).
- Elisa Fromont is responsible (and teaches) for the Master MEEF NSI (M1 & M2) which train new teachers in computer sciences for high school.
- Since September 2020, Sébastien Ferré is the head of Master M1 Miage, and of the EIT international master track in Data Science (about 75 students).
- Since October 2021, Peggy Cellier is responsible of the last year at Computer Science Department at INSA (master 2 level, about 70 students).
- Since September 2019, Tassadit Bouadi is responsible of continuation of studies at IUT of Lannion (computer science department).
- Christine Largouët is head of the computer science educational unit in Institut Agro Rennes Angers (2 engineering schools).

Courses responsibility

- Alexandre Termier is responsible for the following courses at ISTIC (Univ. Rennes): Object Programming (L2 info, elec, maths), AI (M1 info), Data Mining and Visualization (M2 SIF).
- Elisa Fromont is responsible of the Machine Learning course (M2 IL) and teaches AI in M1 Info.
- Luis Galárraga gave a 4h seminar on "Interpretable AI" (M2 MIAGE, Oct 2022).
- At INSA, Peggy Cellier is responsible of four courses: "Databases and web development" (Licence 3 INFO), "Databases" (Licence 3 Math), "Data Mining" (Licence 3) and "Advanced Database and Semantic Web" (Master 2). She also teaches some other courses: "Database" (Licence 2), "Use and functionalities of an operating system" (Licence 3). At master 2 SIF, she teaches in English 4 hours in the data mining course (DMV). In addition she gives a lecture of 2 hours also in master 2 SIF about "Qu'est-ce qu'une thèse, un doctorat, un-e doctorant-e?".
- Sébastien Ferré is responsible of 5 courses at ISTIC: "Basics of Data Analysis with Python" (M1 Miage EIT, in English), "Semantic Web Technologies" (M1 Miage, in English), "Data Mining" (M2 Miage, in English), "Compilers" (M1 info), "Technological Watch" (M1 Miage EIT).

- Romaric Gaudel is responsible for the following course at ISTIC (Univ. Rennes 1): Data analysis and probabilistic modeling (M2 SIF).
- Participation in the module “Case Study in Data Science” with the seminar “Interpretable AI”, M2 EIT DSC, ISTIC, Univ. Rennes I (L. Galárraga, 4h).
- Tassadit Bouadi is responsible for the following courses at IUT of Lannion (Univ. Rennes 1): SAé Creation of a database (BUT1 info) and Exploitation of a database (BUT1 info). And she is co-responsible of SQL and Programming course (BUT2 info).
- Christine Largouët is responsible of the following courses at Institut Agro - Rennes Angers: Databases (L3), Programming in Python (L3), Scientific Programming (M1), Data Management and Machine Learning (M1), Big Data (M2 datascience).

11.2.2 Supervision

Postdocs

- Luis Galárraga co-supervised **Mohit Mihal**, post-doctoral fellow from the Sequel team at Inria Lille in collaboration with Philippe Preux working on “Inspecting and Debugging VQA Systems”. He also collaborates with the **Daniel Hernández**, post-doctoral fellow from Univ. of Stuttgart, Katja Hose from Aalborg University, and Giorgos Flouris and Zubaria Asma from FORTH-ICS in the development of more efficient query rewriting methods to compute how-provenance explanations for SPARQL query results.

PhD. Students

- (defended in 2022) **Johanne Bakalara**, 2018-2022; supervisors: Thomas Guyet, Emmanuel Oger, Olivier Dameron and André Happe; title: Temporal Models of Care Sequences for the Exploration of Medico-administrative Data.
- (defended in 2022) **Samuel Felton**, (PhD, UR1) 2019-2022; supervisors: Éliisa Fromont and Eric Marchand; title: Deep Learning for End-to-end Visual Servoing.
- (defended in 2022) **Camille-Sovanneary Gauthier**, (PhD, CIFRE Vuitton) 2019-2022; supervisors: Romaric Gaudel and Éliisa Fromont, title: Bandit-based Recommender Systems.
- (defended in 2022) **Gregory Martin**, (PhD, CIFRE PSA) 2019-2022 supervisors: Éliisa Fromont, Laurence Rozé, and Alexandre Termier; title: data-driven vehicle relocation in free floating carsharing services with combinatorial optimization.
- (defended in 2022) **Olivier Pelgrin**, (PhD, AAU) 2019-2022; supervisors: Katja Hose and Luis Galárraga; title: Fully-fledged Archiving for RDF Datasets.
- **Josie Signe**, 2020-2022; supervisors: Peggy Cellier, Yannick Le Cozler (Inrae), Véronique Masson and Alexandre Termier, title: Animal Welfare, Characterizing the Diversity between and within Livestock Farming Situations with Data Mining Methods used on Information from Dairy Herd Sensors, ED MathStic.
- **Abderaouf Nassim Amalou** (PhD, UR1/Projects) 2020-2023; supervisors: Éliisa Fromont and Isabelle Puaut; title: Machine Learning for Timing Estimation.
- **Simon Corbillé**, (PhD, UR1) 2019-2023; supervisors: Éliisa Fromont and Eric Anquetil; title: Explainable Deep-learning-based Methods for Children Handwriting Analysis in Education.
- **Victor Guyomard**, 2020-2023; supervisors: Tassadit Bouadi, Thomas Guyet, Françoise Fessant (Orange Labs) and Alexandre Termier, title: Explaining individual decisions made by an AI algorithm.
- **Antonin Voyez**, (PhD, CIFRE Enedis) 2020-2023; supervisors: Éliisa Fromont, Tristan Allard and Gildas Avoine; title: Privacy-preserving Power Consumption Time-series Publishing.

- **Lénaïg Cornanguer**, 2020-2023, supervisors: Christine Largouët, Laurence Rozé and Alexandre Termier; title: Timed Automata Learning, ED MathSTIC
- **Julien Delaunay**, (Inria, ANR) 2020-2023; supervisors: Luis Galárraga and Christine Largouët; title: Automatic Construction of Explanations for AI Models, ED MathSTIC.
- **Maëva Durand**, 2020-2023; supervisors: Christine Largouët and Charlotte Gaillard (INRAE); title: Real-time Integration of Gestating Sow Welfare and Health from Heterogeneous Data for Precision Feeding, ED EGAAL.
- **Hugo Ayats**, 2020-2023; supervisors: Peggy Cellier and Sébastien Ferré, title: De la prédiction à l'automatisation avec une IA explicable et centrée-utilisateur – application à la construction de graphes de connaissances, ED MathStic.
- **Julie Boudebs**, 2021-2024; supervisors: Peggy Cellier and Sébastien Ferré, title: Un assistant en langue naturelle pour interroger le Web sémantique, ED MathStic.
- **Olivier Gauriau**, (Inria, DigitAg, Acta Toulouse) 2021-2024; supervisors: Luis Galárraga, François Brun, Alexandre Termier and David Makowski; title: Numerical Rule Mining for the Prediction of the Dynamics of Crop Diseases.
- **Elodie Germani**, 2021-2024; supervisors: Élixa Fromont and Camille Maumet; title: on representation learning for more robust fMRI data analysis.
- **Gwladys Kelodjou**, 2022-2025; supervisors: Véronique Masson, Laurence Rozé, Alexandre Termier; title: Beyond the oracle: stabilizing the interpretability of machine learning algorithms, ED MathSTIC.
- **Lucie Lepetit**, 2022-2025; supervisors: Peggy Cellier, Bruno Crémilleux and Alexandre Termier; title: Data mining methods for discovering behaviors related to animal well-being in precision farming data, ED MathSTIC.
- **Pierre Maurand**, 2022-2025; supervisors: Tassadit Bouadi, Peggy Cellier, Bruno Crémilleux and Alexandre Termier; title: Tell me your preferences and I will show you what you are interested in, ED MathSTIC.

11.2.3 Juries

- Alexandre Termier was a member of the following PhD juries in 2022: Guillaume Guarino, 9/12 Strasbourg (committee member); Gregory Martin, 15/12 Rennes (co-supervisor)
- Elisa Fromont was a member of the following PhD juries in 2022: Damien Robissout, 10/02 Saint-Etienne (committee member, president); Luxin Zhang, 7/03/2021 Lille (committee member, president); Camille-Sovanneary Gauthier, 17/03, Rennes (co-supervisor); Edward Beeching, 3/05 Lyon (reviewer); Xudong Zhang, 11/05 Paris (committee member, president); Rémi Viola, 24/06 Saint-Etienne (reviewer); Farah Cherfaoui, 11/07 Aix-Marseille (committee member); Baptiste Roziere, 12/07 Paris (committee member); Julia Cohen, 13/07 Lyon (committee member, president); Gaël Aglin, Louvain-la-neuve, Belgium 7/09 (assessor), 19/10 (public defense); Lize Coenen, Leuven, Belgium 16/09 (assessor), 20/10 (public defense); Denis Coquenot, Rouen 29/09 (committee member); Vincent Lequertier, 30/09 Lyon (committee member); Rémy Sun, Paris 17/10 (committee member); Michael Mbouopda, 13/12 Clermont-Ferrand (committee member); Gregory Martin, 15/12 Rennes (co-supervisor); Samuel Felton, 20/12 Rennes (co-supervisor). (HDR) Emmanuelle Becker 14/12 Rennes (committee member, president), Laetitia Chapel, 12/05 Vannes (committee member, president).
- Peggy Cellier was a member of the following PhD juries: Jean Dupuy, 06/05/2022, Université Lyon 2 (reviewer); Nicolas Sourbier, INSA Rennes, 29/09/2022 (committee member); .

- Sébastien Ferré was a member of the following PhD juries in 2022: Salah Boukhetta, 30/08 La Rochelle (committee member, examiner); Clara Delahaye, 15/12 Rennes (committee member, president).
- Christine Largouët was a member of the following PhD juries: Maximilien Cosme (Université de Montpellier), 28/03/2022 (reviewer), Vianey Sicard (Université Bretagne Loire), 28/11/2022 (committee member) and Colin Thomas (Université Paris Saclay) 12/12/2022 (reviewer).

Doctoral advisory comitee (CSID)

- Elisa Fromont was a member of the mid-term evaluation juries of Malik Kazi Aoual (Paris) 7/11/2022; Paul Estano (Rennes) 2/05/2022; Florent Imbert (Rennes) 13/06/2022; Duc Hau (Rennes) 21/05/22 & 9/06/2021; Hasnaa Ouadoudi Belabzioui (Rennes); Célia Wafa AYAD (Paris) 16/09/2022; Michael Franklin MBOUOPDA (Clermont) 16/06/22 & 06/2021; Rita Fermanian & Brandon LeBon (Rennes) 25/05/22 & 29/04/21; Etienne Meunier (Rennes) 31/05/22 & 17/05/2021; Manal HAMZAOUI (Vannes) les 13/06/22 & 2/06/21 & 8/06/20;
- Luis Galárraga was a member of mid-term evaluation committee of the following PhD candidates: Hugo Ayats (Univ. Rennes I, 05/2022) working on relation extraction via concepts of neighbors, Louis Béziaud (Univ. Rennes I/UAQM, 05/2022) working on privacy and fairness in law.
- Peggy Cellier was a member of mid-term evaluation committee of the following PhD candidates: Oumaima El Khettari (Nantes), Triss Jacquot (Caen).
- Sébastien Ferré was a member of the mid-term evaluation juries of Thimotée Neithoffer.
- Christine Largouët was a member of the mid-term evaluation of Baptiste Sorin, 13/12/2022 (INRAE BIOEPAR, Nantes)

11.3 Popularization

11.3.1 Education

- Since September 2018, Tassadit Bouadi is co-responsible of the program **L codent L créent** that raises awareness of digital careers in middle schools. The action is only dedicated to middle school girls.
- Elisa Fromont is jury in the national "trophees-NSI".

11.3.2 Interventions

- Lénaïg Cornanguer participated to the program **L codent L créent** that raises awareness of programming among middle school girls during creative workshops.

12 Scientific production

12.1 Major publications

- [1] V. Bellon Maurel, L. Brossard, F. Garcia, N. Mitton and A. Termier. *Agriculture and Digital Technology: Getting the most out of digital technology to contribute to the transition to sustainable agriculture and food systems*. Jan. 2022, pp. 1–185. DOI: [10.17180/wmkb-ty56-en](https://doi.org/10.17180/wmkb-ty56-en). URL: <https://hal.inrae.fr/hal-03604970>.
- [2] L. Cornanguer, C. Largouët, L. Rozé and A. Termier. 'TAG: Learning Timed Automata from Logs'. In: AAAI 2022 - 36th AAAI Conference on Artificial Intelligence. Virtual, Canada, 22nd Feb. 2022, pp. 1–9. URL: <https://hal.inria.fr/hal-03564455>.

- [3] K. Fauvel, E. Fromont, V. Masson, P. Faverdin and A. Termier. ‘XEM: An explainable-by-design ensemble method for multivariate time series classification’. In: *Data Mining and Knowledge Discovery* 36.3 (11th Feb. 2022), pp. 917–957. DOI: [10.1007/s10618-022-00823-6](https://doi.org/10.1007/s10618-022-00823-6). URL: <https://hal.inria.fr/hal-03599214>.
- [4] K. Fauvel, V. Masson, E. Fromont, P. Faverdin and A. Termier. ‘Towards Sustainable Dairy Management - A Machine Learning Enhanced Method for Estrus Detection’. In: *KDD 2019 - ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 25th SIGKDD Conference on Knowledge Discovery and Data Mining proceedings. Anchorage, United States, Aug. 2019, pp. 1–9. DOI: [10.1145/3292500.3330712](https://doi.org/10.1145/3292500.3330712). URL: <https://hal.archives-ouvertes.fr/hal-02190790>.
- [5] E. Galbrun, P. Cellier, N. Tatti, A. Termier and B. Crémilleux. ‘Mining Periodic Patterns with a MDL Criterion’. In: *European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD)*. Dublin, Ireland, 2018. URL: <https://hal.archives-ouvertes.fr/hal-01951722>.
- [6] C.-S. Gauthier, R. Gaudel and E. Fromont. ‘UniRank: Unimodal Bandit Algorithm for Online Ranking’. In: *ICML 2022 - 39th International Conference on Machine Learning*. Baltimore, United States, 17th July 2022, pp. 1–31. URL: <https://hal.inria.fr/hal-03740981>.
- [7] C.-S. Gauthier, R. Gaudel, E. Fromont and B. A. Lompo. ‘Parametric Graph for Unimodal Ranking Bandit’. In: *ICML 2021 - International Conference on Machine Learning*. Vol. 139. Proceedings of the 38th International Conference on Machine Learning. Virtual, Canada, 2021, pp. 3630–3639. URL: <https://hal.archives-ouvertes.fr/hal-03256621>.
- [8] T. Guyet and R. Quiniou. ‘NegPSpan: efficient extraction of negative sequential patterns with embedding constraints’. In: *Data Mining and Knowledge Discovery* 34 (2020), pp. 563–609. DOI: [10.1007/s10618-019-00672-w](https://doi.org/10.1007/s10618-019-00672-w). URL: <https://hal.inria.fr/hal-03025572>.

12.2 Publications of the year

International journals

- [9] S. Amer-Yahia, A. Bonifati, C. Favre, E. Fromont, N. Labroche, G. Melançon, F. Sèdes, A. Soulet and A. Termier. ‘Diversity and Inclusion Activities in EGC - A 2022 Report’. In: *SIGKDD explorations : newsletter of the Special Interest Group (SIG) on Knowledge Discovery & Data Mining* 24.1 (2nd June 2022), pp. 52–56. DOI: [10.1145/3544903.3544911](https://doi.org/10.1145/3544903.3544911). URL: <https://hal.archives-ouvertes.fr/hal-03878581>.
- [10] K. Fauvel, E. Fromont, V. Masson, P. Faverdin and A. Termier. ‘XEM: An explainable-by-design ensemble method for multivariate time series classification’. In: *Data Mining and Knowledge Discovery* 36.3 (11th Feb. 2022), pp. 917–957. DOI: [10.1007/s10618-022-00823-6](https://doi.org/10.1007/s10618-022-00823-6). URL: <https://hal.inria.fr/hal-03599214>.
- [11] S. Felton, P. Brault, E. Fromont and E. Marchand. ‘Visual Servoing in Autoencoder Latent Space’. In: *IEEE Robotics and Automation Letters* 7.2 (Apr. 2022), pp. 3234–3241. DOI: [10.1109/LRA.2022.3144490](https://doi.org/10.1109/LRA.2022.3144490). URL: <https://hal.inria.fr/hal-03506036>.
- [12] R. Gauthier, C. Largouët, D. Bussièrès, J.-P. Martineau and J.-Y. Dourmad. ‘Precision feeding of lactating sows: implementation and evaluation of a decision support system in farm conditions’. In: *Journal of Animal Science* (2022), pp. 1–40. DOI: [10.1093/jas/skac222](https://doi.org/10.1093/jas/skac222). URL: <https://hal.archives-ouvertes.fr/hal-03699447>.
- [13] O. Krichen, S. Corbillé, E. Anquetil, N. Girard, E. Fromont and P. Nerdeux. ‘Combination of explicit segmentation with Seq2Seq recognition for fine analysis of children handwriting’. In: *International Journal on Document Analysis and Recognition* (7th Nov. 2022). DOI: [10.1007/s10032-022-00409-4](https://doi.org/10.1007/s10032-022-00409-4). URL: <https://hal.archives-ouvertes.fr/hal-03845144>.
- [14] Y. Zhang, S. Destercke, Z. Zhang, T. Bouadi and A. Martin. ‘On computing evidential centroid through conjunctive combination: an impossibility theorem’. In: *IEEE Transactions on Artificial Intelligence* (8th June 2022), pp. 1–10. DOI: [10.1109/TAI.2022.3180973](https://doi.org/10.1109/TAI.2022.3180973). URL: <https://hal.archives-ouvertes.fr/hal-03698839>.

International peer-reviewed conferences

- [15] A. N. Amalou, E. Fromont and I. Puaut. ‘CATREEN : Context-Aware Code Timing Estimation with Stacked Recurrent Networks’. In: ICTAI 2022 - 34th IEEE International Conference on Tools with Artificial Intelligence. Virtually, China: IEEE, 31st Oct. 2022, pp. 1–6. URL: <https://hal.archives-ouvertes.fr/hal-03890057>.
- [16] H. Ayats, P. Cellier and S. Ferré. ‘A Two-Step Approach for Explainable Relation Extraction’. In: IDA 2022 - Symposium on Intelligent Data Analysis. Rennes, France, 20th Apr. 2022, pp. 1–12. URL: <https://hal.inria.fr/hal-03866083>.
- [17] H. Ayats, P. Cellier and S. Ferré. ‘CONNOR: Exploring Similarities in Graphs with Concepts of Neighbors’. In: ETAFCA 2022 - Existing Tools and Applications for Formal Concept Analysis. Tallinn, Estonia, 20th June 2022, pp. 1–6. URL: <https://hal.inria.fr/hal-03866075>.
- [18] L. Cornanguer, C. Largouët, L. Rozé and A. Termier. ‘TAG: Learning Timed Automata from Logs’. In: AAAI 2022 - 36th AAAI Conference on Artificial Intelligence. Virtual, Canada, 22nd Feb. 2022, pp. 1–9. URL: <https://hal.inria.fr/hal-03564455>.
- [19] J. Delaunay, L. Galárraga and C. Largouët. ‘When Should We Use Linear Explanations?’ In: CIKM 2022 - 31st ACM International Conference on Information and Knowledge Management. Atlanta, United States: ACM, 17th Oct. 2022, pp. 355–364. DOI: [10.1145/3511808.3557489](https://doi.org/10.1145/3511808.3557489). URL: <https://hal.inria.fr/hal-03908363>.
- [20] S. Ferré. ‘Exploring the Application of Graph-FCA to the Problem of Knowledge Graph Alignment’. In: CLA 2022 - 16th International Conference on Concept Lattices and Their Applications. Tallinn, Estonia, 20th June 2022, pp. 1–12. URL: <https://hal.inria.fr/hal-03866030>.
- [21] S. Ferré and P. Cellier. ‘Modeling Complex Structures in Graph-FCA: Illustration on Natural Language Syntax’. In: ETAFCA 2022 - Existing Tools and Applications for Formal Concept Analysis. Tallinn, Estonia, 20th June 2022, pp. 1–6. URL: <https://hal.inria.fr/hal-03866048>.
- [22] V. Fokou, P. Cellier, M. Tchuente and A. Termier. ‘QuickFill, QuickMixte : approches par blocs pour la réduction du nombre de programmes en synthèse de programmes: QuickFill, QuickMixte: block approaches for reducing the number of programs in program synthesis’. In: CARI 2022 - Colloque Africain sur la Recherche en Informatique et en Mathématiques Appliquées. Yaoundé, Cameroon, 10th July 2022, pp. 1–10. URL: <https://hal.inria.fr/hal-03717860>.
- [23] R. Gaudel, L. Galárraga, J. Delaunay, L. Rozé and V. Bhargava. ‘s-LIME: Reconciling Locality and Fidelity in Linear Explanations’. In: IDA 2022 - Symposium on Intelligent Data Analysis. Rennes, France, 20th Apr. 2022, pp. 1–13. URL: <https://hal.inria.fr/hal-03741042>.
- [24] C.-S. Gauthier, R. Gaudel and E. Fromont. ‘UniRank: Unimodal Bandit Algorithm for Online Ranking’. In: ICML 2022 - 39th International Conference on Machine Learning. Baltimore, United States, 17th July 2022, pp. 1–31. URL: <https://hal.inria.fr/hal-03740981>.
- [25] V. Guyomard, F. Fessant, T. Guyet, T. Bouadi and A. Termier. ‘VCNet: A self-explaining model for realistic counterfactual generation’. In: ECML PKDD 2022 - European Conference on Machine Learning and Knowledge Discovery in Databases. Grenoble, France, 19th Sept. 2022, pp. 1–16. URL: <https://hal.inria.fr/hal-03899151>.
- [26] I. Harada, K. Fauvel, T. Guyet, V. Masson, A. Termier and P. Faverdin. ‘XPM: An explainable-by-design pattern-based estrus detection approach to improve resource use in dairy farms’. In: *Proceedings of the 36th AAAI Conference on Artificial Intelligence. Workshop. AAAI 2022 - 36th AAAI Conference on Artificial Intelligence*. Vancouver, Canada, 22nd Feb. 2022, pp. 1–10. URL: <https://hal.inria.fr/hal-03483109>.

National peer-reviewed Conferences

- [27] H. Ayats. ‘Construction de Graphes de Connaissance à partir de textes avec une I.A. centrée-utilisateur’. In: *Actes de la 29e Conférence sur le Traitement Automatique des Langues Naturelles. Volume 2 : 24e Rencontres Etudiants Chercheurs en Informatique pour le TAL (RECITAL)*. TALN 2022 - 29e Conférence sur le Traitement Automatique des Langues Naturelles. Avignon, France: ATALA, 2022, pp. 33–46. URL: <https://hal.archives-ouvertes.fr/hal-03704042>.

Conferences without proceedings

- [28] V. Bellon Maurel, L. Brossard, F. Garcia, N. Mitton and A. Termier. ‘The french INRIA- INRAE White Book “Agriculture and digital technologies” - Coupling digitalization and agroecology to design more sustainable and resilient food systems’. In: NARO 2022 - 2nd International Symposium on Food and Health “Global Food Security and Human Health -Toward Transforming the Food System with Advanced Technology-”. virtuel, Japan, 2022. URL: <https://hal.inrae.fr/hal-03900024>.
- [29] R. Gaudel, L. Galárraga, J. Delaunay, L. Roze and V. Bhargava. ‘S-LIME : reconcilier localité et fidélité des explications linéaires’. In: CAP - RFIAP 2022 - Conférence française sur l’apprentissage machine - Reconnaissance de formes, Images, Apprentissage et Perception. Vannes, France, 5th July 2022. URL: <https://hal.inria.fr/hal-03741047>.
- [30] C.-S. Gauthier, R. Gaudel and É. Fromont. ‘UniRank : un algorithme de bandit générique pour l’ordonnancement en ligne’. In: CAP - RFIAP 2022 - Conférence française sur l’apprentissage machine - Reconnaissance de formes, Images, Apprentissage et Perception. Vannes, France, 5th July 2022. URL: <https://hal.inria.fr/hal-03741036>.
- [31] V. Guyomard, F. Fessant, T. Bouadi and T. Guyet. ‘Générer des explications contrefactuelles à l’aide d’un autoencodeur supervisé’. In: Extraction et Gestion des Connaissances, EGC’2022. Extraction et Gestion des Connaissances, EGC’2022. Blois, France, 1st Feb. 2022. URL: <https://hal.archives-ouvertes.fr/hal-03551329>.
- [32] G. Martin, M. Donain, E. Fromont, T. Guns, L. Roze and A. Termier. ‘Optimisation du Positionnement de Voitures en Autopartage basée sur la Prédiction de leur Utilité’. In: Conférence Nationale en Intelligence Artificielle 2022 (CNIA 2022). Actes CNIA 2022. Saint-Etienne, France, 27th June 2022, pp. 1–10. URL: <https://hal.archives-ouvertes.fr/hal-03866192>.
- [33] L. Vallet, F. Brun, O. Gauriau, F. Joudelat, C. Gouwie, G. Malatesta, J. Adrian and F. Maupas. ‘Prédiction des épidémies de cercosporiose de la betterave par des approches d’apprentissage automatique.’ In: VÉGÉPHYL -13ème Conférence internationale sur les maladies des plantes. Orléans, France, 6th Dec. 2022, pp. 1–3. URL: <https://hal.archives-ouvertes.fr/hal-03918080>.
- [34] A. Voyez, T. Allard, G. Avoine, P. Cauchois, E. Fromont and M. Simonin. ‘Membership Inference Attacks on Aggregated Time Series with Linear Programming’. In: SECRYPT 2022 - 19th International Conference on Security and Cryptography. Lisbon, France: SCITEPRESS - Science and Technology Publications, 11th July 2022, pp. 193–204. DOI: [10.5220/0011276100003283](https://doi.org/10.5220/0011276100003283). URL: <https://hal.archives-ouvertes.fr/hal-03726234>.
- [35] H. Zhang, E. Fromont, S. Lefèvre and B. Avignon. ‘Low-cost Multispectral Scene Analysis with Modality Distillation’. In: WACV 2022 - IEEE Winter Conference on Applications of Computer Vision. Waikoloa /Virtual, United States: IEEE, 4th Jan. 2022, pp. 1–10. URL: <https://hal.archives-ouvertes.fr/hal-03491950>.

Scientific books

- [36] V. Bellon Maurel, L. Brossard, F. Garcia, N. Mitton and A. Termier. *Agriculture and Digital Technology: Getting the most out of digital technology to contribute to the transition to sustainable agriculture and food systems*. Jan. 2022, pp. 1–185. DOI: [10.17180/wmkb-ty56-en](https://doi.org/10.17180/wmkb-ty56-en). URL: <https://hal.inrae.fr/hal-03604970>.
- [37] V. Bellon Maurel, L. Brossard, F. Garcia, N. Mitton and A. Termier. *Agriculture et numérique: Tirer le meilleur du numérique pour contribuer à la transition vers des agricultures et des systèmes alimentaires durables*. INRIA, Jan. 2022, pp. 1–195. DOI: [10.17180/wmkb-ty56](https://doi.org/10.17180/wmkb-ty56). URL: <https://hal.inrae.fr/hal-02887646>.

Scientific book chapters

- [38] L. Brossard, V. Bellon Maurel, P. Bisquert, T. Bouadi, F. Garcia, R. Lenain, P. P. Labarthe, E. Lutton, P. Maurel, N. Mitton and A. Termier. ‘Challenges for the future: Providing digital tools to help address the challenges and gaps facing agriculture’. In: *Agriculture and Digital Technology: Getting the most out of digital technology to contribute to the transition to sustainable agriculture and food systems*. White book Inria 6. 2022, pp. 112–134. URL: <https://hal.inrae.fr/hal-03667491>.
- [39] L. Brossard, V. Bellon Maurel, P. Bisquert, T. Bouadi, F. Garcia, R. Lenain, P. P. Labarthe, E. Lutton, P. Maurel, N. Mitton and A. Termier. ‘Les défis de demain: Comment fournir des outils numériques pour aider à répondre aux enjeux et manques en agriculture’. In: *Agriculture et numérique : Tirer le meilleur du numérique pour contribuer à la transition vers des agricultures et des systèmes alimentaires durables*. Livre blanc INRIA 6. INRIA, 2022, pp. 120–143. URL: <https://hal.inrae.fr/hal-03609537>.
- [40] S. Ferré. ‘Conceptual Navigation in Large Knowledge Graphs’. In: *Complex Data Analytics with Formal Concept Analysis*. Springer, 2022, p. 29. URL: <https://hal.inria.fr/hal-03926161>.
- [41] N. Mitton, L. Brossard, T. Bouadi, F. Garcia, R. Gautron, N. Hilgert, D. Ienco, C. Largouët, E. Lutton, V. Masson, R. Martin-Clouaire, M.-L. Mugnier, P. Neveu, P. Preux, H. Raynal, C. Roussey, A. Termier and V. Bellon Maurel. ‘Foundations and state of the art’. In: *Agriculture and Digital Technology: Getting the most out of digital technology to contribute to the transition to sustainable agriculture and food systems*. White book Inria 6. INRIA, 2022, pp. 30–75. URL: <https://hal.inrae.fr/hal-03609470>.

Doctoral dissertations and habilitation theses

- [42] C.-S. Gauthier. ‘List recommendations with multi-armed bandits’. Université Rennes 1, 17th Mar. 2022. URL: <https://theses.hal.science/tel-03854857>.

Reports & preprints

- [43] A. N. Amalou, É. Fromont and I. Puaut. *CATREEN: Context-Aware Code Timing Estimation with Stacked Recurrent Networks*. 13th Sept. 2022. URL: <https://hal.inria.fr/hal-03776508>.
- [44] E. Germani, E. Fromont and C. Maumet. *On the benefits of self-taught learning for brain decoding*. 2022. URL: <https://hal.inria.fr/hal-03769993>.
- [45] A. Voyez, T. Allard, G. Avoine, P. Cauchois, E. Fromont and M. Simonin. *Unique in the Smart Grid - The Privacy Cost of Fine-Grained Electrical Consumption Data*. 9th Nov. 2022. URL: <https://hal.archives-ouvertes.fr/hal-03833605>.

12.3 Cited publications

- [46] J. Castells, P.-D. Mohammad, L. Galárraga, G. Méndez, M. Ortiz-Rojas and A. Jiménez. ‘A Student-oriented Tool to Support Course Selection in Academic Counseling Sessions’. In: *Proceedings of the Workshop on Adoption, Adaptation and Pilots of Learning Analytics in Under-represented Regions co-located with the 15th European Conference on Technology Enhanced Learning 2020*. Virtual Event, Germany, Sept. 2020. URL: <https://hal.inria.fr/hal-03084671>.
- [47] M. Feurer, A. Klein, K. Eggenberger, J. Springenberg, M. Blum and F. Hutter. ‘Efficient and Robust Automated Machine Learning’. In: *Advances in Neural Information Processing Systems 28*. Ed. by C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama and R. Garnett. Curran Associates, Inc., 2015, pp. 2962–2970. URL: <http://papers.nips.cc/paper/5872-efficient-and-robust-automated-machine-learning.pdf>.
- [48] G. Mendez, L. Galárraga and K. Chiluita. ‘Showing Academic Performance Predictions during Term Planning: Effects on Students’ Decisions, Behaviors, and Preferences’. In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. CHI ’21. Yokohama, Japan: Association for Computing Machinery, 2021. DOI: [10.1145/3411764.3445718](https://doi.org/10.1145/3411764.3445718). URL: <https://doi.org/10.1145/3411764.3445718>.