UMR IRISA

Activity Report 2023

Team LACODAM

Large scale Collaborative Data Mining

Joint team with Centre Inria de l'Université de Rennes

D7 – Data and Knowledge Management

Université de Rennes | rennes | INSA INSTITUT NATIONAL DES SCIENCES APPLIQUÉES RENNES | cnrs | Inria | Inserm | CentraleSupélec | IMT Atlantique Bretagne-Pays de la Loire École Mines-Télécom | Université Bretagne Sud ubs

Contents

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, Researcher]

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 Ferré [UNIV RENNES I, Professor, HDR]
- Elisa Fromont [UNIV RENNES I, Professor, HDR]
- Romaric Gaudel [UNIV RENNES I, Associate Professor, HDR]
- Christine Largouët [L'INSTITUT AGRO, Associate Professor, HDR]
- Véronique Masson [UNIV RENNES I, Associate Professor]
- Laurence Rozé [INSA RENNES, Associate Professor]

PhD Students

- H. Ambre Ayats [UNIV RENNES I, until Sep 2023]
- Julie Boudebs [UNIV RENNES I]
- Isseïnie Calviac [UNIV RENNES I, from Sep 2023]
- Simon Corbille [UNIV RENNES I, until Mar 2023, with Intuidoc Team]
- Lénaïg Cornanguer [INRIA, until Nov 2023]
- Julien Delaunay [INRIA]
- Olivier Gauriau [ACTA Asso, CIFRE]
- Elodie Germani [UNIV RENNES I, with EMPENN Team]
- Victor Guyomard [ORANGE LABS, CIFRE, until Nov 2023]
- Yasmine Hachani [INRIA, from Oct 2023, with SAIRPICO Team]
- Gwladys Kelodjou [UNIV RENNES I]
- Charbel Gaspard Kindji [ORANGE LABS, CIFRE]
- Lucie Lepetit [INRIA]
- Dimitri Lereverend [INRIA, from Sep 2023, with WIDE Team]
- Pierre Maurand [INSA RENNES]
- Paul Sevellec [STELLANTIS, CIFRE, from Oct 2023]
- Antonin Voyez [ENEDIS, CIFRE, until May 2023, with SPICY Team]

Technical Staff

- Louis Bonneau De Beaufort [L'INSTITUT AGRO, Engineer, Research Engineer]
- Maiwenn Fleig [UNIV RENNES I, Engineer, from Feb 2023 until Jun 2023]

Interns and Apprentices

- Aymeric Behaegel [ENS DE LYON, Intern, until May 2023]
- Isseïnie Calviac [INRIA, Intern, until Jul 2023]
- Thibault Chanus [UNIV RENNES I, Intern, from Oct 2023]
- Niels Cobat [UNIV RENNES I, Intern, from May 2023 until Aug 2023]
- Raphael Giraud [ENS PARIS-SACLAY, Intern, from Jun 2023 until Jul 2023]
- Selim Gmati [UNIV RENNES I, Intern, until Feb 2023]
- Sergiu Mocanu [INRIA, Intern, from May 2023 until Aug 2023]
- Alexandra Padonou [UNIV RENNES I, Intern, from May 2023 until Aug 2023]

Administrative Assistant

- Gaelle Tworkowski [INRIA]

Visiting Scientists

- Vanessa Fokou [Université Strasbourg, from Jun 2023 until Jun 2023]
- Gonzalo Mendez [ESPOL, Ecuador, from Jul 2023 until Jul 2023]

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 digital 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 in April 2022 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 strengthens 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.

- **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, funded 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.

- **Heterogeneous tabular data generation with deep generative models** Tabular data generation is paramount when dealing with privacy-sensitive data and with missing values, which are frequent cases in the real (industrial) world and particularly at Orange. It is also used for data augmentation, a pre-processing step often needed when training data-hungry deep learning models (for example to detect anomalies in networks, study customer profiles, ...). The CIFRE PhD of Charbel Kindji, funded by Orange, is concerned with this application. We study methods to tackle this problem when the tabular data are heterogeneous (numerical and symbolic) and when new tables should be generated from scratch based on a human prompt.

- **Counterfactual explanations over multivariate time series**. Very complex machine learning models (that are called-black boxes) are often used in critical applications (e.g. self-driving cars). To comply with EU regulations and better understand their systems, many companies, and in particular Stellantis, are interested in developing skills in "explainable AI", a domain which aims at bringing back the human in the decision loop that involves a black box model. The CIFRE PhD of Paul Sevellec, funded by Stellantis, is concerned with this application. We study the particular case of counterfactual explanations on the challenging context of multivariate time-series. This problem is related to the generation of new data that fulfills some human requirements.

4.2 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.

- **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.

- **Deep learning-based analysis of the early development of bovine embryos from videomicroscopy**. The PhD of Yasmine Hachani (collaboration with team Sairpico and INRAE) focuses on designing deep learning methods for the comparison and classification of videos of embryos produced in vitro (PIV). These automatic methods are eagerly awaited by biologists in order to broaden the potential of fundamental and applied research in this field, and to help improve results and reproductive performance in breeding. The problem posed is multifaceted. First of all, the images acquired by microscopy are complex in nature: they are low-contrast, noisy, contain transparency effects, and movements are difficult to characterize. The categorization of in vitro fertilized embryos, in terms of the quality of their development, is based on a continuum of classes, rather than distinct ones. Furthermore, the need is to obtain reliable classification at the earliest possible stage, i.e. 3 days post-gamete contact, from a video of 300 images, with images acquired every 15 minutes. Finally, while classification can be supervised, we have only a limited amount of data (a few hundred videos) for deep learning purposes, especially as class characterization can only be achieved by observing a video in its entirety.

4.3 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 CTI [1] 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's handwriting is, at an early stage of learning, approximate and includes deformed letters. This is a joint effort between the LACODAM and ShaDoc (IRISA) teams.

4.4 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 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 operations[2]. 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.

[1] Centro de Tecnologías de Información
[2] Processes where there is some sort of data difference

Work trips. While the sanitary crisis had drastically cut the number of work trips of the team, recent years have seen an increase in the physical participation in 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 and most of them using trains (and not planes). 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 recent PhD research topics (e.g., the theses of Simon Corbillé, and Simon Felton) concern deep learning technologies, and the important place of eXplanaible AI in our research program have made our team highly reliant on Igrida (notably with the PhD of Julien Delaunay and Victor Guyomard). While the discontinuation of Igrida services and the transition towards Grid'5000 and Jean Zay has reduced our access to easily available computation resources (making it harder to perform experiments during the transition and requiring to learn new ways of operating), it can be said that it has a positive effect on energy consumption, as we are now using national infrastructures that benefit from even better sharing between users than Igrida (which was already heavily used).

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 is 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 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 received the "prix de thèse EGC 2023" for his PhD thesis on "Mining tractable sets of graph patterns with the minimum description length principle", co-supervised by Sébastien Ferré and Peggy Cellier and defended on 23/11/2021.

Sébastien Ferré won a Novelty Prize at the ARCathon 2023 challenge, where he also ranked 6th out of 24 active participants as team MADIL. The Abstraction and Reasoning Corpus (ARC) is a benchmark introduced by François Chollet (AI researcher at Google) in 2019 to measure AI skill acquisition and foster research toward human-level AI.

Simon Corbillé (supervised by Elisa Fromont and Eric Anquetil) obtained the best poster award for his work on "Precise Segmentation for Children Handwriting Analysis by Combining Multiple Deep Models with Online Knowledge" at ICDAR 2023.

Maëva Durand (supervised by Christine Largouet and Charlotte Gaillard (INRAE)) won the PhD prize of the Association Française de Zootechnie (AFZ), for her PhD "Alimentation sur mesure et estimation du bien-être des truies gestantes à partir de données hétérogènes". This PhD is funded by the DigitAg project.

6.2 Other highlights

Romaric Gaudel defended his HDR in February on the subject "Recommender Systems: Online Learning and Ranking"

Elisa Fromont and Alexandre Termier were involved in the project proposal for the CMA AI training project (6 M euros) which was awarded in February for the Université of Rennes. Elisa Fromont is now the scientific leader of this 5 years project.

7 New software, platforms, open data

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

7.1.2 Dexteris

Keywords: Data Exploration, Querying, Interactive method, JSon

Functional Description: Dexteris is a low-code tool for data exploration and transformation. It works as an interactive data-oriented query builder with JSONiq as the target query language. It uses JSON as the pivot data format but it can read from and write to a few other formats: text, CSV, and RDF/Turtle (to be extended to other formats).

Dexteris is very expressive as JSONiq is Turing-complete, and supports a varied set of data processing features: - reading JSON files, and CSV as JSON (one object per row, one field per column), - string processing (split, replace, match, ...), - arithmetics, comparison, and logics, - accessing and creating JSON data structures, i- terations, grouping, filtering, aggregates and ordering (FLWOR operators), - local function definitions.

The built JSONiq programs are high-level, declarative, and concise. Under-progress results are given at every step so that users can keep focused on their data and on the transformations they want to apply.

URL: http://www.irisa.fr/LIS/ferre/dexteris/

Publication: hal-04186117v1

Contact: Sebastien Ferre

7.1.3 skm

Name: scikit-mine

Keywords: Artificial intelligence, Data mining, Pattern discovery, Sequential patterns

Functional Description: The library offers several algorithms for extracting a reasonable-sized set of patterns for different types of data (itemsets, sequences, graphs).

URL: https://scikit-mine.github.io/scikit-mine/

Contact: Peggy Cellier

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: H. Ambre Ayats, Peggy Cellier, Lénaïg Cornanguer, Sébastien Ferré, 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.

Concepts of Neighbors and their Application to Instance-based Learning on Relational Data [20]. Knowledge graphs and other forms of relational data have become a widespread kind of data, and powerful methods to analyze and learn from them are needed. Formal Concept Analysis (FCA) is a mathematical framework for the analysis of symbolic datasets, which has been extended to graphs and relational data, like Graph-FCA. It encompasses various tasks such as pattern mining or machine learning, but its application generally relies on the computation of a concept lattice whose size can be exponential with the number of instances. We propose to follow an instance-based approach where the learning effort is delayed until a new instance comes in, and an inference task is set. This is the approach adopted in k-Nearest Neighbors, and this relies on a distance between instances. We define a conceptual distance based on FCA concepts, and from there the notion of concepts of neighbors, which can be used as a basis for instance-based reasoning. Those definitions are given for both classical FCA and Graph-FCA. We provide efficient algorithms for computing concepts of neighbors, and we demonstrate their inference capabilities by presenting three different applications: query relaxation, knowledge graph completion, and relation extraction.

Data Mining-Based Techniques for Software Fault Localization [50]. This chapter illustrates the basic concepts of fault localization using a data mining technique. It utilizes the Trityp program to illustrate the general method. Formal concept analysis and association rule are two well-known methods for symbolic data mining. In their original inception, they both consider data in the form of an object-attribute table. In their original inception, they both consider data in the form of an object-attribute table. The chapter considers a debugging process in which a program is tested against different test cases. Two attributes, PASS and FAIL, represent the issue of the test case. The chapter extends the analysis of data mining for fault localization for the multiple fault situations. It addresses how data mining can be further applied to fault localization for GUI components. Unlike traditional software, GUI test cases are usually event sequences, and each individual event has a unique corresponding event handler.

Scikit-mine: A pattern mining library in Python [58]. A poster presenting the scikit-mine library, a python library for pattern mining. This library proposes an Open Source implementation of recent MDL-based pattern mining algorithms, that focuses on the ease of use of these algorithms.

Persistence-Based Discretization for Learning Discrete Event Systems from Time Series [47]. To get a good understanding of a dynamical system, it is convenient to have an interpretable and versatile model of it. Timed discrete event systems are a kind of model that respond to these requirements. However, such models can be inferred from timestamped event sequences but not directly from numerical data. To solve this problem, a discretization step must be done to identify events or symbols in the time series. Persist is a discretization method that intends to create persisting symbols by using a score called persistence score. This allows to mitigate the risk of undesirable symbol changes that would lead to a too complex model. After the study of the persistence score, we point out that it tends to favor excessive cases making it miss interesting persisting symbols. To correct this behavior, we replace the metric used in the persistence score, the Kullback-Leibler divergence, with the Wasserstein distance. Experiments show that the improved persistence score enhances Persist's ability to capture the information of the original time series and that it makes it better suited for discrete event systems learning.

Using the MDL principle for automating unique abstraction and reasoning tasks (ARC) from a few examples [35]. The ARC (Abstraction and Reasoning Corpus) challenge has been proposed to push AI research towards more generalization capability rather than ever more performance. It is a collection of unique tasks about generating colored grids, specified by a few examples only. We propose object-centered models analogous to the natural programs produced by humans. The MDL (Minimum Description Length) principle is exploited for an efficient search in the vast model space. We obtain encouraging results with a class of simple models: various tasks are solved and the learned models are close to natural programs.

Sky-signatures: detecting and characterizing recurrent behavior in sequential data [25]. The signature approach considers a sequence of itemsets, and given a number k it returns a segmentation of the sequence in k segments such that the number of items occurring in all segments is maximized. The limitation of this approach is that it requires to manually set k, and thus fixes the temporal granularity at which the data is analyzed. We propose the sky-signature model that removes this requirement, and allows us to examine the results at multiple levels of granularity, while keeping a compact output. We also propose efficient algorithms to mine sky-signatures, as well as an experimental validation with real data both from the retail domain and from natural language processing (political speeches).

8.1.2 Graph-FCA

Participants: H. Ambre Ayats, Peggy Cellier, Sébastien Ferré.

Some otherwise presented documents also contribute to this research domain: [20] .

Graph-FCA meets Pattern Structures [34]. A number of extensions have been proposed for Formal Concept Analysis (FCA). Among them, Pattern Structures (PS) bring complex descriptions on objects, as an extension to sets of binary attributes; while Graph-FCA brings n-ary relationships between objects, as well as n-ary concepts. We have introduced a novel extension named Graph-PS that combines the benefits of PS and Graph-FCA. In conceptual terms, Graph-PS can be seen as the *meet* of PS and Graph-FCA, seen as sub-concepts of FCA. We have demonstrated how it can be applied to RDFS graphs, handling hierarchies of classes and properties, and patterns on literals such as numbers and dates.

8.1.3 Semantic Web

Participants: H. Ambre Ayats, Peggy Cellier, Sébastien Ferré, Luis Galárraga.

Some of previously presented documents also contribute to this research domain: [34] .

Language Models as Controlled Natural Language Semantic Parsers for Knowledge Graph Question Answering [41], in collaboration with Prof. Jens Lehmann (Amazon, TU Dresden). We have proposed the use of controlled natural language as a target for knowledge graph question answering (KGQA) semantic parsing via language models as opposed to using formal query languages directly. Controlled natural languages are close to (human) natural languages, but can be unambiguously translated into a formal language such as SPARQL. Our research hypothesis is that the pre-training of large language models (LLMs) on vast amounts of textual data leads to the ability to parse into controlled natural language for KGQA with limited training data requirements. We devise an LLM-specific approach for semantic parsing to study this hypothesis. To conduct our study, we created a dataset that allows the comparison of one formal and two different controlled natural languages. Our analysis shows that training data requirements are indeed substantially reduced when using controlled natural languages, which is of relevance since collecting and maintaining high-quality KGQA semantic parsing training data is very expensive and time-consuming.

Scaling Large RDF Archives To Very Long Histories [44]. In recent years, research in RDF archiving has gained traction due to the ever-growing nature of semantic data and the emergence of community-maintained knowledge bases. Several solutions have been proposed to manage the history of large RDF graphs, including approaches based on independent copies, time-based indexes, and change-based schemes. In particular, aggregated changesets have been shown to be relatively efficient at handling very large datasets. However, ingestion time can still become prohibitive as the revision history increases. To tackle this challenge, we propose a hybrid storage approach based on aggregated changesets, snapshots, and multiple delta chains. We evaluate different snapshot creation strategies on the BEAR benchmark for RDF archives, and show that our techniques can speed up ingestion time up to two orders of magnitude while keeping competitive performance for version materialization and delta queries. This allows us to support revision histories of lengths that are beyond reach with existing approaches.

GLENDA: Querying over RDF Archives with SPARQL [43]. The dynamicity of semantic data has propelled the research on RDF Archiving, i.e., the task of storing and making the full history of a large RDF dataset accessible. That said, existing archiving techniques fail to scale when confronted to very large RDF archives and complex SPARQL queries. In this demonstration, we showcase GLENDA, a system capable of running full SPARQL 1.1 compliant queries over large RDF archives. We achieve this through a multi-snapshot change-based storage architecture that we interface using the Comunica query engine. Thanks to this integration we demonstrate that fast SPARQL query processing over multiple versions is possible. Moreover our demonstration provides different statistics about the history of RDF datasets. This provides insights about the evolution dynamics of the data.

The Need for Better RDF Archiving Benchmarks [49]. The advancements and popularity of Semantic Web technologies in the last decades have led to an exponential adoption and availability of Web-accessible datasets. While most solutions consider such datasets to be static, they often evolve over time. Hence, efficient archiving solutions are needed to meet the users' and maintainers' needs. While some solutions to these challenges already exist, standardized benchmarks are needed to systematically test the different capabilities of existing solutions and identify their limitations. Unfortunately, the development of new benchmarks has not kept pace with the evolution of RDF archiving systems. In this paper, we therefore identify the current state of the art in RDF archiving benchmarks and discuss to what degree such benchmarks reflect the current needs of real-world use cases and their requirements. Through this empirical assessment, we highlight the need for the development of more advanced and comprehensive benchmarks that align with the evolving landscape of RDF archiving.

Concepts of Neighbors and their Application to Instance-based Learning on Relational Data [20]. Knowledge graphs and other forms of relational data have become a widespread kind of data, and powerful methods to analyze and learn from them are needed. Formal Concept Analysis (FCA) is a mathematical framework for the analysis of symbolic datasets, which has been extended to graphs and relational data, like Graph-FCA. It encompasses various tasks such as pattern mining or machine learning, but its application generally relies on the computation of a concept lattice whose size can be exponential with the number of instances. We propose to follow an instance-based approach where the learning effort is delayed until a new instance comes in, and an inference task is set. This is the approach adopted in k-Nearest Neighbors, and this relies on a distance between instances. We define a conceptual distance based on FCA concepts, and from there the notion of concepts of neighbors, which can be used as a basis for instance-based reasoning. Those definitions are given for both classical FCA and Graph-FCA. We provide efficient algorithms for computing concepts of neighbors, and we demonstrate their inference capabilities by presenting three different applications: query relaxation, knowledge graph completion, and relation extraction.

8.1.4 Data Wrangling

Participants: Sébastien Ferré.

Dexteris: Data Exploration and Transformation with a Guided Query Builder Approach [33]. Data exploration and transformation remain a challenging prerequisite to the application of data analysis methods. The desired transformations are often ad-hoc so that existing end-user tools may not suffice, and plain programming may be necessary. We propose a guided query builder approach to reconcile expressivity and usability, i.e. to support the exploration of data, and the design of ad-hoc transformations, through data-user interaction only. This approach is available online as a client-side web application, named Dexteris. Its strengths and weaknesses are evaluated on a representative use case, and compared to plain programming and ChatGPT-assisted programming.

8.2 Interpretable Machine Learning

Participants: Tassadit Bouadi, Simon Corbillé, Julien Delaunay, Élisa Fromont, Luis Galárraga, Victor Guyomard, Christine Largouët, Alexandre Termier.

"Honey, Tell Me What's Wrong", Global Explainability of NLP Models through Cooperative Generation [45]. The ubiquity of machine learning has highlighted the importance of explainability algorithms. Among these, model-agnostic methods generate artificial examples by slightly modifying original data and then observing changes in the model's decision-making on these artificial examples. However, such methods require initial examples and provide explanations only for the decisions based on these examples. To address these issues, we propose Therapy, the first model-agnostic explainability method for language models that does not require input data. This method generates texts that follow the distribution learned by the classifier to be explained through cooperative generation. Not depending on initial examples allows for applicability in situations where no data is available (e.g., for privacy reasons) and offers explanations on the global functioning of the model instead of multiple local explanations, thus providing an overview of the model's workings. Our experiments show that even without input data, Therapy provides instructive insights into the text features used by the classifier, which are competitive with those provided by methods using data.

Precise Segmentation for Children Handwriting Analysis by Combining Multiple Deep Models with Online Knowledge [31]. We present a strategy, called Seq2Seg, to reach both precise and accurate recognition and segmentation for children handwritten words. Reaching such high performance for

both tasks is necessary to give personalized feedback to children who are learning how to write. The first contribution is to combine the predictions of an accurate Seq2Seq model with the predictions of a R-CNN object detector. The second one is to refine the bounding box predictions provided by the detector with a segmentation lattice computed from the online signal. An ablation study shows that both contributions are relevant, and their combination is efficient enough for immediate feedback and achieves state-of-the-art results even compared to more informed systems.

Adaptation of AI Explanations to Users' Roles [48]. Surrogate explanations approximate a complex model by training a simpler model over an interpretable space. Among these simpler models, we identify three kinds of surrogate methods: (a) feature-attribution, (b) example-based, and (c) rule-based explanations. Each surrogate approximates the complex model differently, and we hypothesise that this can impact how users interpret the explanation. Despite the numerous calls for introducing explanations for all, no prior work has compared the impact of these surrogates on specific user roles (e.g., domain expert, developer). In this article, we outline a study design to assess the impact of these three surrogate techniques across different user roles.

Effects of Locality and Rule Language on Explanations for Knowledge Graph Embeddings [36]. Knowledge graphs (KGs) are key tools in many AI-related tasks such as reasoning or question answering. This has, in turn, propelled research in link prediction in KGs, the task of predicting missing relationships from the available knowledge. Solutions based on KG embeddings have shown promising results in this matter. On the downside, these approaches are usually unable to explain their predictions. While some works have proposed to compute post-hoc rule explanations for embedding-based link predictors, these efforts have mostly resorted to rules with unbounded atoms, e.g., bornIn(x,y)->residence(x,y), learned on a global scope, i.e., the entire KG. None of these works has considered the impact of rules with bounded atoms such as nationality(x,England)->speaks(x,English), or the impact of learning from regions of the KG, i.e., local scopes. We therefore study the effects of these factors on the quality of rule-based explanations for embedding-based link predictors. Our results suggest that more specific rules and local scopes can improve the accuracy of the explanations. Moreover, these rules can provide further insights about the inner-workings of KG embeddings for link prediction.

Visualizing How-Provenance Explanations for SPARQL Queries [37]. Knowledge graphs (KGs) are vast collections of machine-readable information, usually modeled in RDF and queried with SPARQL. KGs have opened the door to a plethora of applications such as Web search or smart assistants that query and process the knowledge contained in those KGs. An important, but often disregarded, aspect of querying KGs is query provenance: explanations of the data sources and transformations that made a query result possible. In this article we demonstrate, through a Web application, the capabilities of SPARQLprov, an engine-agnostic method that annotates query results with how-provenance annotations. To this end, SPARQLprov resorts to query rewriting techniques, which make it applicable to already deployed SPARQL endpoints. We describe the principles behind SPARQLprov and discuss perspectives on visualizing how-provenance explanations for SPARQL queries.

Interactive Visualization of Counterfactual Explanations for Tabular Data [39], [46]. In this paper, we present an interactive visualization tool that exhibits counterfactual explanations to explain model decisions. Each individual sample is assessed to identify the set of changes needed to flip the output of the model. These explanations aim to provide end-users with personalized actionable insights with which to understand automated decisions. An interactive method is also provided so that users can explore various solutions. The functionality of the tool is demonstrated by its application to a customer retention dataset. The tool is compatible with any counterfactual explanation generator and decision model for tabular data.

Generating robust counterfactual explanations [38]. Counterfactual explanations have become a mainstay of the XAI field. This particularly intuitive statement allows the user to understand what small but necessary changes would have to be made to a given situation in order to change a model prediction. The quality of a counterfactual depends on several criteria: realism, actionability, validity, robustness, etc.

In this paper, we are interested in the notion of robustness of a counterfactual. More precisely, we focus on robustness to counterfactual input changes. This form of robustness is particularly challenging as it involves a trade-off between the robustness of the counterfactual and the proximity with the example to explain. We propose a new framework, CROCO, that generates robust counterfactuals while managing effectively this trade-off, and guarantees the user a minimal robustness. An empirical evaluation on tabular datasets confirms the relevance and effectiveness of our approach.

Impressions and Strategies of Academic Advisors When Using a Grade Prediction Tool During Term Planning [42]. Academic advising brings numerous benefits to the mission of Higher Education Institutions. One of the main actors is the advisors who support students in defining appropriate academic roadmaps. One central and challenging duty of academic advisors is course recommendation for term planning. This task requires both knowledge of the study programs and a thorough analysis of the students' unique circumstances. If we add limited time and a large student population, the task becomes overwhelming. As a result, an important body of research has sought to expedite term planning via data-oriented decision-support tools. While the impact of such tools on students has been extensively studied, the advisors' perspective remains largely unexplored. We contribute to filling this gap by studying how a grade prediction tool shapes advisors' course recommendation strategies. Our observations suggest that while the advisors' usual strategies tend to prevail, their recommendations are mostly affected by the advisee's historical performance. CCS Concepts: • Human-centered computing → Empirical studies in visualization.

8.3 Real World AI

8.3.1 Computer Vision and Robotics

Participants: Simon Corbillé, Samuel Felton, Élisa Fromont, Elodie Germani.

Some of previously presented documents also contribute to this research domain: [31] .

Deep metric learning for visual servoing: when pose and image meet in latent space [32]. We propose a new visual servoing method that controls a robot's motion in a latent space. We aim to extract the best properties of two previously proposed servoing methods: we seek to obtain the accuracy of photometric methods such as Direct Visual Servoing (DVS), as well as the behavior and convergence of pose-based visual servoing (PBVS). Photometric methods suffer from limited convergence area due to a highly non-linear cost function, while PBVS requires estimating the pose of the camera which may introduce some noise and incurs a loss of accuracy. Our approach relies on shaping (with metric learning) a latent space, in which the representations of camera poses and the embeddings of their respective images are tied together. By leveraging the multimodal aspect of this shared space, our control law minimizes the difference between latent image representations thanks to information obtained from a set of pose embeddings. Experiments in simulation and on a robot validate the strength of our approach, showing that the sought-out benefits are effectively found.

On the benefits of self-taught learning for brain decoding [26]. Context. 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. Results. 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 samples available both for pre-training and finetuning the models and on the complexity of the targeted downstream task. Conclusion. The pre-trained model improves the classification performance and displays more generalizable features, less sensitive to individual differences.

Uncovering communities of pipelines in the task-fMRI analytical space [55], [60], [61]. Functional magnetic resonance imaging analytical workflows are highly flexible with no definite consensus on how to choose a pipeline. While methods have been developed to explore this analytical space, there is still a lack of understanding of the relationships between the different pipelines. We use community detection algorithms to explore the pipeline space and assess its stability across different contexts. We show that there are subsets of pipelines that give similar results, especially those sharing specific parameters (e.g., number of motion regressors, software packages, etc.), with relative stability across groups of participants. By visualizing the differences between these subsets, we describe the effect of pipeline parameters and derive general relationships in the analytical space.

The HCP multi-pipeline dataset: an opportunity to investigate analytical variability in fMRI data analysis [56]. Results of functional Magnetic Resonance Imaging (fMRI) studies can be impacted by many sources of variability including differences due to: the sampling of the participants, differences in acquisition protocols and material but also due to different analytical choices in the processing of the fMRI data. While variability across participants or across acquisition instruments has been extensively studied in the neuroimaging literature the root causes of analytical variability remain an open question. Here, we share the *HCP multi-pipeline dataset*, including the resulting statistic maps for 24 typical fMRI pipelines on 1,080 participants of the HCP-Young Adults dataset. We share both individual and group results - for 1,000 groups of 50 participants - over 5 motor contrasts. We hope that this large dataset covering a wide range of analysis conditions will provide new opportunities to study analytical variability in fMRI.

8.3.2 Agriculture

Participants: Christine Largouët.

Real-time combination of observed growth and feed intake performance with performance simulated by InraPorc® to apply precision feeding to growing pigs [57]. Real-time combination of observed growth and feed intake performance with performance simulated by InraPorc® to apply precision feeding to growing pigs Precision feeding (PF) of growing pigs requires methods for real-time analysis of performance and prediction of nutritional requirements. Two calculation methods for reducing nutrient intake were evaluated. Individual daily kinetics of body weight (BW) and feed intake (FI) of 285 pigs, reared from 81 to 156 days of age (ad libitum feeding) were used. The PF1 approach (from the Feeda-Gene project) used the Holt-Winters and MARS methods to predict individual daily FI and BW, respectively. The standardised digestible lysine (dLys) requirement was calculated daily from the predicted performance using the factorial method. The PF2 method used 2200 virtual pigs whose performance was simulated using InraPorc®. By comparing FI and BW dynamics of real and virtual pigs, the 10 closest virtual pigs were selected daily for each real pig. Individual daily performance and expected dLys requirements were obtained by averaging the InraPorc® data of these 10 virtual pigs. PF was then simulated for each real pig. For each method, the blend proportions of two diets (A and B, 9.7 MJ net energy (NE)/kg, crude protein: 16.9% and 9.3%, dLys: 1.0 and 0.4 g/MJ NE, respectively) were calculated daily to achieve calculated requirements. Nitrogen (N) and dLys intake and N excretion were calculated individually. A two-phase (2-P) feeding strategy was also simulated (A:B = 83:17 until 65 kg PV, 50:50 afterwards). Compared to 2-P, N and dLys intake and N excretion were reduced by, respectively, 6.7%, 9.7% and 11.9% with PF1 and by 9.2%, 13.3% and 16.3% with PF2. The PF2 method provided better day-to-day stability of performance predictions, leading to a more regular decrease in N and dLys intakes during growth. The potential of this new method needs to be confirmed under field conditions.

A dataset to study group-housed sows' individual behaviours and production responses to different short-term events [21]. The relational database SOWELL was created to better understand the behaviour and individual responses of gestating sows facing different short-term events induced: a competitive situation for feed, hot and cold thermal conditions, a sound event, an enrichment (straw, ropes and bags

available) and an impoverishment (no straw, no objects) of the pen. The data were collected on 102 cross-bred sows equipped with activity sensors, group-housed in video-recorded pens (16–18 sows per pen), with access to automatons. Feeding and drinking behaviours were extracted from the electronic feeders and drinkers' recordings. Social behaviours, physical activities and locations in the pen were recorded thanks to manual video analysis labelling at the individual scale. Accelerometer fixed on the sows' ears also recorded individual physical activities. The physical activity was also determined at a group scale by automatic video analysis using deep learning techniques. BWs, back fat thickness, and body condition (cleanliness, body damages) were recorded weekly during the whole gestation. Last gestation room data regarding environmental conditions (temperature, humidity, noise level) were recorded using automatic sensors. The database can fulfil different research purposes, namely sows' nutrition for example to better calculate the energy requirements regarding environmental factors, or also on welfare or health during gestation by providing indicators.

Prediction of daily nutritional requirements of gestating sows based on their behaviour and machine learning methods [59]. Background and Objectives Precision feeding aims to define the right feeding strategy according to individual's nutrient requirements, in order to improve health and reduce feed cost. Usually, the nutrient requirements of gestating sows are provided by a mechanistic nutritional model requiring input data such as age and body status. This paper proposes to predict the daily nutritional requirements, with only the data measured by sensors. According to various digital farm configurations, we explore and evaluate Machine Learning (ML) methods to predict nutrient requirements of gestating sows. Material and Methods Behavioural data of gestating sows are extracted from sensors data collected on 73 sows from parities 1 to 9. Their nutrient requirements concerned metabolisable energy (ME, in MJ/d) and standard ileal digestible lysine (SID Lys, in g/ d). Various digital farm configurations are proposed, from low-cost to more expensive equipments (electronic feeder and drinker, connected weight scale, accelerometers and video analysis software), producing various data at different levels of detail on sow behavior. Nine ML algorithms were trained on these 23 scenarios to predict daily energy and lysine for each sow. Results proposed by the ML algorithms are compared with outputs given by the nutritional model InraPorc. Results Using a Random Forest algorithm, the RMSE were lower with data feeder alone (1.22 MJ/d for ME and 0.53 g/d for SID Lys, 2.4 and 4.02% of mean absolute error respectively) compared those obtained with combined data from feeders and accelerometers (1.01 MJ/d and 0.29 g/d, 1.9 and 2.1%). The inclusion of the sows' characteristics reduced the RMSE, on average, by 20% for ME and by 35% for Lys. Discussion and Conclusion This study highlights that daily requirements of gestating sows can be predicted accurately thanks to behavioural data provided by sensors. It paves the way to propose simpler solutions for the application of precision feeding on farms.

Prediction of the daily nutrient requirements of gestating sows based on sensor data and machine-learning algorithms [23]. Precision feeding is a strategy for supplying an amount and composition of feed as close that are as possible to each animal's nutrient requirements, with the aim of reducing feed costs and environmental losses. Usually, the nutrient requirements of gestating sows are provided by a nutrition model that requires input data such as sow and herd characteristics, but also an estimation of future farrowing performances. New sensors and automatons, such as automatic feeders and drinkers, have been developed on pig farms over the last decade, and have produced large amounts of data. This study evaluated machine-learning methods for predicting the daily nutrient requirements of gestating sows, based only on sensor data, according to various configurations of digital farms. The data of 73 gestating sows was recorded using sensors such as electronic feeders and drinker stations, connected weight scales, accelerometers, and cameras. Nine machine-learning algorithms were trained on various dataset scenarios according to different digital farm configurations (one or two sensors), in order to predict the daily metabolizable energy and standardized ileal digestible lysine requirements for each sow. The prediction results were compared to those predicted by the InraPorc model, a mechanistic model for the precision feeding of gestating sows. The scenario predictions were also evaluated with or without the housing conditions and sow characteristics at artificial insemination usually integrated into the InraPorc model. Adding housing and sow characteristics to sensor data improved the mean average percentage error by 5.58% for lysine and by 2.22% for energy. The higher correlation coefficient values for lysine (0.99) and for energy (0.95) were obtained for scenarios involving an automatic feeder system (daily

duration and number of visits with or without consumption) only. The scenarios including an automatic feeder combined with another sensor gave good performance results. For the scenarios using sow and housing characteristics and automatic feeder only, the root mean square error was lower with Gradient Tree Boosting (0.91 MJ/d for energy and 0.08 g/d for lysine) compared with those obtained using linear regression (2.75 MJ/d and 1.07 g/d). The results of this study show that the daily nutrient requirements of gestating sows can be predicted accurately using data provided by sensors and machine-learning methods. It paves the way to simpler solutions for precision feeding.

Estimation of gestating sows' welfare status based on machine learning methods and behavioral data [22]. Estimating the welfare status at an individual level on the farm is a current issue to improve livestock animal monitoring. New technologies showed opportunities to analyze livestock behavior with machine learning and sensors. The aim of the study was to estimate some components of the welfare status of gestating sows based on machine learning methods and behavioral data. The dataset used was a combination of individual and group measures of behavior (activity, social and feeding behaviors). A clustering method was used to estimate the welfare status of 69 sows (housed in four groups) during different periods (sum of 2 days per week) of gestation (between 6 and 10 periods, depending on the group). Three clusters were identified and labelled (scapegoat, gentle and aggressive). Environmental conditions and the sows' health influenced the proportion of sows in each cluster, contrary to the characteristics of the sow (age, body weight or body condition). The results also confirmed the importance of group behavior on the welfare of each individual. A decision tree was learned and used to classify the sows into the three categories of welfare issued from the clustering step. This classification relied on data obtained from an automatic feeder and automated video analysis, achieving an accuracy rate exceeding 72%. This study showed the potential of an automatic decision support system to categorize welfare based on the behavior of each gestating sow and the group of sows.

8.3.3 Machine Learning on Sequences

Participants: Abderaouf N Amalou, Élisa Fromont.

CAWET: Context-Aware Worst-Case Execution Time Estimation Using Transformers [28]. This paper presents CAWET, a hybrid worst-case program timing estimation technique. CAWET identifies the longest execution path using static techniques, whereas the worst-case execution time (WCET) of basic blocks is predicted using an advanced language processing technique called Transformer-XL. By employing Transformers-XL in CAWET, the execution context formed by previously executed basic blocks is taken into account, allowing for consideration of the microarchitecture of the processor pipeline without explicit modeling. Through a series of experiments on the TacleBench benchmarks, using different target processors (Arm Cortex M4, M7, and A53), our method is demonstrated to never underestimate WCETs and is shown to be less pessimistic than its competitors.

8.3.4 Software Engineering

Participants: Peggy Cellier, Sébastien Ferré.

Data Mining-Based Techniques for Software Fault Localization [50]. This book chapter illustrates the basic concepts of fault localization using a data mining technique. It utilizes the Trityp program to illustrate the general method. Formal concept analysis and association rule are two well-known methods for symbolic data mining. In their original inception, they both consider data in the form of an object-attribute table. In their original inception, they both consider data in the form of an object-attribute table. The chapter considers a debugging process in which a program is tested against different test cases. Two attributes, PASS and FAIL, represent the issue of the test case. The chapter extends the analysis of data

mining for fault localization for the multiple fault situations. It addresses how data mining can be further applied to fault localization for GUI components. Unlike traditional software, GUI test cases are usually event sequences, and each individual event has a unique corresponding event handler.

9 Bilateral contracts and grants with industry

9.1 Bilateral contracts with industry

- **ORANGE - Univ. Rennes**

  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 regard 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.

- **Stellantis - Univ. Rennes**

  Participants: Elisa Fromont, Romaric Gaudel, Laurence Rozé, Paul Sévellec.

  Contract amount: 70k€ + Phd Salary

  Context. This project is a collaboration with Stellantis and focuses on the development of interpretable machine learning models for multivariate time series data. Utilizing a range of sensors integrated within vehicles, these models are designed to make real-time decisions. Providing drivers with clear explanations of these decisions is a key aspect. We specifically concentrate on counterfactual explanations, which not only clarify why a particular decision was made but also illustrate how alternative scenarios might have led to different outcomes.

  Objective. Current approaches providing counterfactual explanations for time series models are limited to univariate time series. In this project, we aim to develop approaches to handle multivariate time series, which requires capturing the correlations between the series.

  Additional remarks. This is the doctoral contract for the PhD of Paul Sévellec (Thèse CIFRE).

- **Enedis - Univ. Rennes**

  Participants: Elisa Fromont, Antonin Voyez.

Contract amount: 45k€ + Phd Salary

Context. 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.

Objective. We are interested in privacy-preserving data-sharing. We study the uniqueness of large-scale real-life fine-grained electrical consumption time-series, the potential privacy threats, and their mitigation.

Additional remarks. This is the doctoral contract for the PhD of Antonin Voyez (Thèse CIFRE).

- **ORANGE - Univ. Rennes**

Participants: Elisa Fromont, Charbel Kindji.

Contract amount: 45k€ + Phd Salary

Context. Tabular data generation is paramount when dealing with privacy-sensitive data and with missing values, which are frequent cases in the real (industrial) world and particularly at Orange. It is also used for data augmentation, a pre-processing step often needed when training data-hungry deep learning models (for example to detect anomalies in networks, study customer profiles, ...).

Objective. We study methods to tackle heterogeneous tabular data generation with deep generative models. We are particularly interested in problems where the tabular data are heterogeneous (numerical and symbolic) and when new tables should be generated from scratch based on a human prompt.

Additional remarks. This is the doctoral contract for the PhD of Charbel Kindji (Thèse CIFRE).

10 Partnerships and cooperations

10.1 International research visitors

10.1.1 Visits of international scientists

Other international visits to the team

Status Researcher

Institution of origin: ESPOL - Escuela Superior Politécnica del Litoral

Country: Ecuador

Dates: 03-07 July

Context of the visit: The visit was done in the context of the ongoing collaboration between Gonzalo Méndez and the LACODAM team concerning two projects: (i) the study of the effects of AI-based recommendations for a course planning on users (students and advisor); (ii) novel visualization techniques for pattern-aided models that predict the incidence of crop diseases.

Mobility program/type of mobility: Research visit (financed by the visitor's home institution)

10.1.2 Visits to international teams

Sabbatical programme

Research stays abroad Elodie Germani (PhD student supervised by Elisa Fromont with EMPENN) has spent 3 months in Canada at Concordia University (Montreal) with a Mitacs Globalink Research Award (GRA) with Centre National de la Recherche Scientifique (CNRS) on the project "Improving rs-fMRI-derived biomarkers of Parkinson's Disease".

10.2 European initiatives

10.2.1 H2020 projects

Elisa Fromont, Alexandre Termier and Luis Galárraga are all members (within Inria) of the project H2020 ICT-48 TAILOR "Foundations of Trustworthy AI - Integrating Reasoning, Learning and Optimization". Elisa Fromont is responsible for Task 3.7 and 3.8 (roadmap and synergies with industry) in WP3.

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. The project ended in June 2023.

- **#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, preserve 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 Largouet, 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 led 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 (15Keuros / year for 5 years). 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, desiging 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 was 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.

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 an explanation that is both understandable by humans and faithful among a universe of possible explanations. We focus on local explanations, i.e., when we want to explain the answer of a black-box model for a given use case, which we call the "target instance". We argue that the choice of the best explanation depends on the (i) data, namely the model, the explanation technique and

the target instance, etc., and (ii) the recipients of the explanations. Hence our research is focused on two main questions: "What makes an explanation suitable (interpretable and faithful) for a particular instance and model?" and "What is the effect of the different AI-based explanation techniques and visual representations on users' comprehension and trust?". Answering these questions will help us understand and automate the selection of a particular explanation style based on the use case. Our ultimate goal is to produce a suite of algorithms that will compute suitable explanations for ML algorithms based on our insights of what is interpretable. User studies on different explanation settings (methods and visual representations) will allow us to characterize the features of explanations that make them acceptable (i.e., understandable and trustworthy) by users.

- **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.

11 Dissemination

11.1 Promoting scientific activities

11.1.1 Scientific events: organisation

As part of the scientific animation of the DKM (D7) research department at IRISA, Elisa Fromont (head of the department) co-organises monthly seminars which have featured, in 2023: Damien Eveillard, Frederic Jurie, Colin de la Higuera, Sihem Amer-Yahia, Meghyn Bienvenu, Hendrik Blockeel, Aurélien Bellet.

General chair, scientific chair Romaric Gaudel was co-program chair of CAP 2023 (the French conference on Machine Learning) in Strasbourg.

Member of the steering committees

- Peggy Cellier is member of the steering committee of the international conference ECML PKDD.

11.1.2 Scientific events: selection

Member of the editorial boards

- Peggy Cellier is member of the editorial board of ICFCA.
- Sébastien Ferré is member of the editorial board of ICFCA.

Chair of conference program committees

- Luis Galarraga Del Prado, Tassadit Bouadi: Organization of the AIMLAI (Advances in Interpretable Machine Learning and Artificial Intelligence) workshop co-located with the ECML/PKDD conference that took place in Turin on the week of September 18-22, 2023.

Member of the conference program committees

- Tassadit Bouadi: IDA'23, ECAI'23, ECMLPKDD'23
- Peggy Cellier: BDA'23 (Demonstration), EGC'23, ICCS'23, ICFCA'23, IDA'23, ECAI'23, ECMLPKDD'23
- Elisa Fromont: KDD'23, IDA'23, ECMLPKDD'23, IJCAI'23
- Sebastien Ferre: ECAI, ESWC, IDA, ICFCA, FQAS, EGC
- Luis Galarraga Del Prado: ECAI'23, IJCAI'23, ESWC'23, ISWC'23

- Alexandre Termier: KDD'23, SDM'23, IDA'23, AIMLAI'23 (workshop)
- Rozé Laurence:AIMLAI'23 (workshop)
- Christine Largouët:APIA'23 (french conference)
- Romaric Gaudel: NeurIPS'23, AAAI'24, CAp'23, AIMLAI'23

Reviewer

- Luis Galarraga Del Prado: The Web Conference 2023
- Christine Largouët: KDD 2023

11.1.3 Journal

Member of the editorial boards

- Elisa Fromont: Co-Specialty Chief Editor (with Andrea Passerini) of Frontiers in Artificial Intelligence specialty Machine Learning and Artificial Intelligence .
- Luis Galarraga Del Prado: Guest Co-Editor (with Miguel Couceiro) of the EURO Journal on Operational Research - special volume on Fair and Explainable Systems
- Alexandre Termier: Member of the editorial board of the Data Mining and Knowledge Discovery (DMKD) journal.

Reviewer - reviewing activities

- Sebastien Ferre: Semantic Web, Int. J. Approximate Reasoning
- Alexandre Termier: Computer and Electronics in Agriculture, Neurocomputing, Machine Learning Journal
- Christine Largouët: Computer and Electronics in Agriculture
- Romaric Gaudel: Transactions on Machine Learning Research

11.1.4 Invited talks

Elisa Fromont:

- 7/12/2023: Invited speaker (in French) about "Intelligence Artificielle, Menaces ou Opportunités ?" Imagine Summit, France.
- 6/12/2023: Actress (plaintiff's expert) in a mock trial of Artificial Intelligence organized by CCI-Ille et Vilaine, Rennes France.
- 14/11/2023: Invited speaker (in French) for a "Conférence sur l'I.A.", médiathèque Noyal-sur Vilaine, France.
- 25/10/2023: Invited speaker (in French) for Confiance.ai: Lessons learned from the HyAIAI ("Hybrid Approaches from Interpretable AI") project. Online seminar.
- 14/07/2023: Training session (in French) on artificial intelligence for Rennes city councillors.
- 23/06/2023: Invited talk on "Explainable Time Series Classification", GADP Data Science Incubator Outreach at Stellantis, online.
- 14/06/2023: Online tutorial on "Explainable Time Series Classification" at ODSC, London.
- 13/06/2023: Pannelist "Chat GPT, la nouvelle révolution numérique " at Inbound Marketing France, Rennes.

- 25/05/2023: Invited talk (in French) for high school students "Intelligence Artificielle de quoi parle-t-on ?", Trophées NSI (finale regionale). Slides here.
- 18/01/2023: Training session for Masters on "Explainable Time Series Classification", Lille (Virtual).

Sebastien Ferre:

- 13/04/2023: Invited speaker (in French) about "Tackling the Abstraction and Reasoning Corpus (ARC) with Object-centric Models and the MDL Principle", GDR IA seminar, online.
- 7/12/2023: Invited speaker (in French) about "Exploring the Application of Graph-FCA to the Problem of Knowledge Graph Alignment", Journées Redescriptions et Graphes, Caen.

Alexandre Termier:

- 10/05/2023: Invited speaker (in French) about "L'IA à la ferme: moissoner les données pour quels fruits ?", Académie d'Agriculture de France, Paris.
- 23/05/2023: Invited speaker (in French) about "Livre blanc Agriculture et Numérique INRAE/Inria", Rendez-vous Franco-Belge de l'Agritech, Gembloux Campus, Belgium (organized by French Embassy in Belgium).
- 24/08/2023: Invited speaker (in French) about "L'IA à la ferme: moissoner les données pour quels fruits ?", Annual congress of agriculture chambers of the Atlantic Arc (AC3A), Bruz.

Peggy Cellier:

- 31/05/2023: Invited poster about "Scikit-mine: A pattern mining library in Python", Spring workshop on Mining and Learning (SMiLe) 2023, Hertogenbosch (Netherland).

Christine Largouët:

- 27/03/2023: Invited speaker on "Timed Automata Learning", Verification Seminar, IRIF, Paris.

Romaric Gaudel:

- 06/09/2023: Invited speaker on "Bandits manchots : agir et apprendre en même temps", Seminar of Computer Science Department of ENS Rennes, Rennes.

11.1.5 Scientific expertise

- **Tassadit Bouadi**: Member of the working group of Axis 2 'Research and Innovation Program' within the IRIS-E program at the University of Rennes
- **Peggy Cellier**: Member of the folowing recruiting committees (2): IUT Annecy/LISTIC, ISTIC/IRISA (3 positions for one committee). Also spare member for the IUT Lannion MCF recruiting comittee.
- **Elisa Fromont**: In 2023, I was a member of the scientific council of INS2I, of SSFAM (the "société savante d'apprentissage machine"), of AFIA ("Association Française pour l'Intelligence Artificielle" for the machine learning track), of the GDR RADIA ("Raisonnement, Apprentissage et Décision en IA") and of the "Pôle de compétitivité" Images & Réseaux in Rennes. I am also on the executive board of the Labex COMINLABS ("Data, AI and Robotics" track). In addition, I was part of the folowing recruiting committees (5): Leuphana University of Lüneburg (Germany), Lannion MCF-MCF1166 (presidente), INSA Rennes MCF, CPJ IA U-Rennes, Repyramidage Université de Rennes 2.
- **Alexandre Termier**: Spare member for the Lannion MCF-MCF1166 recruiting comittee.
- **Christine Largouët**: Member of the CSTP of the PEPR Agroecology and ICT.

11.1.6 Research administration

- **Peggy Cellier**: Peggy Cellier is in charge of the Phd students of the IRISA lab (commission personnel each month, etc). She is elected at the "Conseil de coposante" of the Computer Science departement of INSA Rennes. She is also a member of "Conseil de l'école doctorale MATISSE".

11.2 Teaching - Supervision - Juries

11.2.1 Teaching

Apart from Luis Galarraga Del Prado (research scientist) and Gaelle Tworkowski (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

- **Véronique Masson** is the head of the L3 studies in Computer Science at University of Rennes
- **Alexandre Termier** is co-head of Master 2 SIF (Science Informatique - research master in Computer Science) at University of Rennes, with Bertrand Coüasnon (INSA Rennes).
- **Sebastien Ferre** is the head of Master M1 Miage, and of the EIT international master track in Data Science (about 75 students).
- **Peggy Cellier** is the head of the last year at Computer Science Department at INSA (master 2 level, about 70 students).
- **Tassadit Bouadi** was head of continuation of studies at IUT of Lannion (computer science department), until July 2023. Since September 2023, she is co-head (with **Romaric Gaudel**) of the future Master M1 and M2 Artificial Intelligence at ISTIC, University of Rennes.
- **Christine Largouet** is head of the computer science educational unit at Institut Agro Rennes Angers (2 engineering schools). Since septembre 2023, she is co-head of the new master M1 and M2 E2C (Water, Energy and Climate, climate change mitigation and adaptation) at Institut Agro Rennes Angers.

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 "Deep Lerning for Vision" (DLV) course (M2 SIF), the Machine Learning course (M2 IL) and teaches AI in M1 Info and L2 Info.
- **Luis Galarraga Del Prado** (i) taught 6h within the course "Semantic Web" (by Sébastien Ferré M1 MIAGE ISTIC, Mar 2023); (ii) was responsible of 42h of teaching (22 TP + 20 TD) for the course "Java Programming" (INSA, Licence 1 INFO)
- **Peggy Cellier** is responsible of four courses at INSA Rennes: "Databases and web development" (Licence 3 INFO), "Databases" (Licence 3 Math), "Data Analysis and Data Mining" (Licence 3 INFO) and "Advanced Database and Semantic Web" (Master 2). She also teaches some other courses: "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 ?".
- **Sebastien Ferre** 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 courses at ISTIC (Univ. Rennes): "discover AI" (L2), "Machine Learning" (M1 SIF) Data analysis and probabilistic modeling (M2 SIF), a course on recommender systems (M2 Miage & IET), a course on information retrieval and natural language processing (M2 Miage).
- **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). Since September 2023, she joined ISTIC, and is responsible of AI (M1 Info) course.
- **Christine Largouet** is responsible of the following courses at Institut Agro - Rennes Angers: Databases (L2 and L3), Programming in Python (L3), Scientific Progamming (M1), Data Management and Machine Learning (M1), Artificial Intelligence (M2 E2C - Water Energy and Climate).
- **Laurence Rozé** is responsible of the following courses at INSA Rennes : prolog (L3), mobile programming (L3,M1).

Other responsibilities

- **Peggy Cellier** is in charge of the APC (Approche par compétences) development for the Computer Science Department. She is also part of the IDPE (Ingénieur diplômé par l'état) committee of INSA Rennes. She also represents INSA Rennes in the CMA (Compétence et Métier d'Avenir) IA TIAre.
- **Laurence Rozé** is in charge of the communication at the computer science departement at INSA Rennes.

11.2.2 Supervision

PhD. Students

- (defended in 2023) **Abderaouf Nassim Amalou** (PhD, UR1/Projects) 2020-2023; supervisors: Elisa Fromont and Isabelle Puaut (PACAP); title: Machine Learning for Timing Estimation, ED Matisse.
- (defended in 2023) **Simon Corbille**, (PhD, UR1) 2019-2023; supervisors: Elisa Fromont and Eric Anquetil; title: Explainable Deep-learning-based Methods for Children Handwriting Analysis in Education, ED Matisse.
- (defended in 2023) **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, ED Matisse.
- (defended in 2023) **Antonin Voyez**, (PhD, CIFRE Enedis) 2020-2023; supervisors: Elisa Fromont, Tristan Allard and Gildas Avoine; title: Privacy-preserving Power Consumption Time-series Publishing, ED Matisse.
- (defended in 2023) **Lenaig Cornanguer**, 2020-2023, supervisors: Christine Largouet, Laurence Rozé and Alexandre Termier; title: Timed Automata Learning, ED Matisse.
- (defended in 2023) **Julien Delaunay**, (Inria, ANR) 2020-2023; supervisors: Luis Galarraga Del Prado and Christine Largouet; title: Automatic Construction of Explanations for AI Models, ED Matisse.
- (defended in 2023) **Maëva Durand**, 2020-2023; supervisors: Christine Largouet and Charlotte Gaillard (INRAE); title: Real-time Integration of Gestating Sow Welfare and Health from Heterogeneous Data for Precision Feeding, ED EGAAL.
- (defended in 2023) **H. Ambre Ayats**, 2020-2023; supervisors: Peggy Cellier and Sebastien Ferre, title: De la prédiction à l'automatisation avec une IA explicable et centrée-utilisateur – application à la construction de graphes de connaissances, ED Matisse.

- **Julie Boudebs**, 2021-2024; supervisors: Peggy Cellier and Sebastien Ferre, title: Un assistant en langue naturelle pour interroger le Web sémantique, ED Matisse.
- **Olivier Gauriau**, (Inria, DigitAg, Acta Toulouse) 2021-2024; supervisors: Luis Galarraga Del Prado, François Brun, Alexandre Termier and David Makowski; title: Numerical Rule Mining for the Prediction of the Dynamics of Crop Diseases, ED Matisse.
- **Elodie Germani**, 2021-2024; supervisors: Elisa Fromont and Camille Maumet (EMPENN); title: on representation learning for more robust FMRI data analysis, ED Matisse.
- **Gwladys Kelodjou**, 2022-2025; supervisors: Véronique Masson, Laurence Rozé, Alexandre Termier; title: Beyond the oracle: stabilizing the interpretability of machine learning algorithms, ED Matisse.
- **Charbel Kindji**, 2022-2025; supervisors: Elisa Fromont and Tanguy Urvoy (OrangeLabs); title: Architectures connexionnistes pour la génération de données tabulaires, ED Matisse.
- **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 Matisse.
- **Pierre Maurand**, 2022-2025; supervisors: Tassadit Bouadi, Peggy Cellier, Bruno Crémilleux (GREYC) and Alexandre Termier; title: Tell me your preferences and I will show you what you are interested in, ED Matisse.
- **Vanessa Fokou**, 2022-2025; supervisors: Florence Le Ber and Xavier Dolques (Univ. Strasbourg), Sebastien Ferre, Peggy Cellier; title: Comparison and cooperation of different Formal Concept Analysis approaches for relational data, Univ. Strasbourg.
- **Isseinie Calviac**, 2023-2026; supervisors: Luis Galarraga Del Prado, Alexandre Termier; title: How-Provenance Polynomials for Efficient and Greener Rule Mining, ED Matisse (financed by an ENS scholarship).
- **Yasmine Hachani**, 2023-2026; supervisors: Patrick Bouthémy (SAIRPICO), Elisa Fromont; title: Analyse par apprentissage profond de la dynamique du développement précoce des embryons bovins à partir de vidéomicroscopie, ED Matisse.
- **Paul Sevellec**, 2023-2026; supervisors: Matteo Sammarco (Stellantis), Elisa Fromont, Romaric Gaudel, Laurence Rozé ; title: Explications de séries temporelles multivariées par contrefactuels, ED Matisse.
- **Dimitri Lereverend**, 2023-2026; supervisors: Davide Frey (WIDE), Romaric Gaudel ; title: Privacy Preserving Decentralized Through Model Fragmentation, ED Matisse.

Internships

- **Selim Gmati** (4th year Engineer Internship) ; supervisor: Laurence Rozé; title: Etude des modèles de diffusion pour la génération d'explication contractuelle de classification de séries temporelles.
- **Isseinie Calviac** (M2) ; supervisors: Luis Galarraga Del Prado, Alexandre Termier; title: How-Provenance Polynomials for Efficient and Greener Rule Mining.
- **Yasmine Hachani** (M2); supervisors: Elisa Fromont, Patrick Bouthémy (SAIRPICO) ; title: Comparaison et classification de vidéos par réseaux neuronaux récurrents : application au potentiel de survie d'embryons bovins.
- **Alexandra Padonou** (M1); supervisors: Peggy Cellier, Sebastien Ferre; title: Evaluation of the capabilities and performance of LLMs for the question-answering of knowledge graphs.
- **Sergiu Mocanu** (M1); supervisors: Alexandre Termier, Sebastien Ferre, Peggy Cellier, Romaric Gaudel ; title: Utilisation de ChatGPT pour la résolution de problèmes identiques à la synthèse de programmes.

- **Niels Cobat** (M1); supervisors: Romaric Gaudel, Damien Hardy (PACAP); title: Estimation précise du temps d'impression 3D par machine learning

- **Raphael Giraud** (M1); supervisors: Romaric Gaudel and François Schwarzentruber (LOGICA); title: Local Search for Cominatorial Bandit Algorithms.

- **Thibault Chanus** (M1); supervisors: Elisa Fromont; title: Image-to-Image Transition Using Diffusion Models in Functional MRI.

Engineer

- **Maiwenn Fleig** (Engineer); supervisors: Elisa Fromont, Elodie Germani ; title: Conception et développement de prototype d'analyses de données d'imageries médicales.

11.2.3 PHD & HDR Juries

- **Tassadit Bouadi** was a member of the following juries in 2023: Victor Guyomard, 23/11 University of Rennes (PhD, co-supervisor).

- **Peggy Cellier** was a member of the following PhD juries in 2023: Maëlle Moranges, 18/01 Univ. Claude Bernard Lyon 1 (PhD, reviewer); Jérémy Richard, 22/05 Univ. de La Rochelle (PhD, examinatrice); Youcef Remil, 06/10 INSA Lyon (PhD, examiner); Albeiro Espinal, 11/12 IMT Atlatique (PhD, examiner); Véronne Yepmo, 20/12 Univ. de Rennes (PhD, examiner), Ambre Ayats Univ. de Rennes (PhD, examiner, co-supervisor).

- **Elisa Fromont** (24, 45% local juries): Antoine Guillaume, 4/01 Orléans (committee member, president); Victor Connes, Nantes 5/01 (committee member, president); William Piat, 06/03 Caen (committee member); Georgios Zervakis, 8/03 Nancy (committee member, president); Corentin Dancette 31/03 Paris (reviewer); Luc Lesoil, 17/04 Rennes (committee member, president); Zed Lee, 21/04 & 24/11 Stockholm, Sweden online(reviewer); Mathieu Chambe, 16/06, Rennes (committee member, president); Simon Corbillé, 28/06, Rennes (co-supervisor); Antonin Voyez, 11/07, Rennes (co-supervisor); Arthur Clavière, 17/07 Toulouse online (committee member); Christophe Genevey Metat, 28/09, Rennes (committee member, president); Nedeljko Radulovic, 29/09, Paris online (committee member); Clément Lalanne, 4/10 Lyon (committee member, president); Julien Ferry, 9/10, Toulouse online (committee member); Victor Epain, 27/11, Rennes (committee member, president); Cyril Li, St-Etienne online, 29/11 (committee member); Etienne Meunier, 4/12, Rennes (committee member, president); Nassim Amalou, 12/12, Rennes (co-supervisor); Thibault Maho, 14/12, Rennes (committee member, president); Benjamin Marais, 18/12, Caen online (committee member); Antonin Deschemps, 19/12 Rennes (committee member); Julien Delaunay, 20/12 Rennes (committee member). (HDR) Romaric Gaudel, 2/02 Rennes (committee member, president).

- **Sebastien Ferre** was a member of the following PhD juries in 2023: Lucas Simonne, 2/2 Univ. Paris-Saclay (rapporteur); Armand Boschin, 21/4 Institut Polytechnique de Paris (rapporteur); Nacira Abbas, 13/10 Univ. de Lorraine (examinateur, président).

- **Luis Galarraga Del Prado** was a member of the following PhD juries in 2023: Luca Veryin, 30/3 INSA Lyon (examiner); Armand Boschin, 21/4 Institut Polytechnique de Paris (examiner); Armita Khajeh Nassiri, 12/07 Université Paris-Saclay (examiner); Nacira Abbas, 13/10 Univ. de Lorraine (examiner); Julien Aimonier-Davat, 11/12 Nantes Université; Louis Beziaud, 6/12 Université de Rennes I (examiner); Julien Delaunay, 20/12 Université de Rennes I (co-supervisor).

- **Alexandre Termier** was a member of the following juries in 2023: Roberto Inderdonato, 23/06 University of Montpellier (HDR, reviewer, not present at defense); Dominique Li, 13/07 University of Tours / IUT of Blois (HDR, reviewer); Youcef Remil, 6/10 INSA Lyon (PhD, reviewer); Nassim Bouarour, 13/12 University of Grenoble Alpes (PhD, reviewer); Lenaig Cornanguer, 2/11 University of Rennes (PhD, examiner, supervisor); Victor Guyomard, 23/11 University of Rennes (PhD, examiner, supervisor).

- **Christine Largouët** was a member of the following juries: Rôlin Gabriel Rasoanaivo, 12/05 University of Toulouse (PhD, examiner), Maeva Durand, 23/10 Institut Agro Rennes Angers (PhD, examiner, supervisor); Lenaig Cornanguer, 2/11 University of Rennes (PhD, examiner, supervisor);

11.2.4 Doctoral advisory comitee (CSID)

- **Peggy Cellier** was a member of the mid-term evaluation committee of Oumaima El Khettari (Université de Nantes); Syrine Hidri (Université de Rennes, IETR) and Jules Berry (INSA Rennes, IRMAR).
- **Sebastien Ferre** was a member of the mid-term evaluation juries of Thimotée Neithoffer.
- **Luis Galarraga Del Prado** was a member of the mid-term evaluation committee of Ataollah Kamal (INSA Lyon) and Sacha Corbugy (Université de Namur)
- **Alexandre Termier** was a member of the mid-term evaluation committee of Valentin Guien (University of Clermont-Ferrand), Armel Soubiega (University of Clermont-Ferrand), Lise Morice (University of Rennes), Erwan Vincent (University of Rennes)
- **Elisa Fromont** was a member of the mid-term evaluation committee of Ewan MOREL-CORLU (Rennes) 14/09/2023; Bruno Michelot (Lyon) 05/09/23; Loïc Eyango (Rennes) 28/03/2023; Irina Proskurina (Lyon) 19/04/23; Ricky Walsh (Rennes) 11/05/23; Paul Estano (Rennes) 9/05/2023 & 2/05/2022; Florent Imbert (Rennes) 5/06/23 & 13/06/2022; Duc Hau (Rennes) 14/06/23 & 21/05/22 & 9/06/2021 and Hasnaa Ouadoudi Belabzioui (Rennes) 15/05/23.
- **Laurence Rozé** was a member of the mid-term evaluation juries of Nolwenn Pinczon du sel.
- **Christine Largouët** was a member of the mid-term evaluation juries of Baptiste Sorin (BIOEPAR, INRAE).

11.3 Popularization

11.3.1 Interventions

- **Luis Galarraga Del Prado** participated (2h) at the tutorial on *eXplainable Graph ML* (in collaboration with Megha Khosla from TU Delft) organized within the AIMLAI workshop that took place at the ECML/PKDD 2023 conference (Sept 18-22).
- **Tassadit Bouadi** is co-organizer of the project *L Codent, L Créent Rennes*, since 2018.

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. URL: https://hal.inrae.fr/hal-03604970.

[2] P. Cellier, M. Ducassé, S. Ferré, O. Ridoux and W. E. Wong. 'Data Mining-Based Techniques for Software Fault Localization'. In: *Handbook of Software Fault Localization*. 1. Wiley, 20th Apr. 2023, Chapitre 7. DOI: 10.1002/9781119880929.ch7. URL: https://hal.science/hal-04403506.

[3] S. Corbillé, E. Anquetil and E. Fromont. 'Precise Segmentation for Children Handwriting Analysis by Combining Multiple Deep Models with Online Knowledge'. In: ICDAR 2023 - 17th International Conference on Document Analysis and Recognition. San José, United States, 21st Aug. 2023, pp. 1–18. URL: https://hal.science/hal-04142592.

[4] 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.

[5] M. Durand, C. Largouët, L. Bonneau de Beaufort, J.-Y. Dourmad and C. Gaillard. 'Prediction of the daily nutrient requirements of gestating sows based on sensor data and machine-learning algorithms'. In: *Journal of Animal Science* 101 (2023), skad337. DOI: 10.1093/jas/skad337. URL: https://hal.inrae.fr/hal-04235874.

[6] 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. URL: https://hal.inria.fr/hal-03599214.

[7] 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. URL: https://hal.archives-ouvertes.fr/hal-02190790.

[8] S. Felton, É. Fromont and E. Marchand. 'Deep metric learning for visual servoing: when pose and image meet in latent space'. In: ICRA 2023 - IEEE International Conference on Robotics and Automation. London, United Kingdom: IEEE, 29th May 2023, pp. 741–747. DOI: 10.1109/ICRA48891.2023.10160963. URL: https://inria.hal.science/hal-04003126.

[9] L. Galárraga, D. Hernández, A. Katim and K. Hose. 'Visualizing How-Provenance Explanations for SPARQL Queries'. In: WWW 2023 - ACM International World Wide Web Conference. Austin, United States: ACM, 2023, pp. 212–216. DOI: 10.1145/3543873.3587350. URL: https://inria.hal.science/hal-04386268.

[10] 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.

[11] 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.

[12] 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.

[13] C. Gautrais, P. Cellier, T. Guyet, R. Quiniou and A. Termier. 'Sky-signatures: detecting and characterizing recurrent behavior in sequential data'. In: *Data Mining and Knowledge Discovery* (29th Aug. 2023). DOI: 10.1007/s10618-023-00949-1. URL: https://hal.science/hal-04401641.

[14] E. Germani, E. Fromont and C. Maumet. 'On the benefits of self-taught learning for brain decoding'. In: *GigaScience* 12 (3rd May 2023), pp. 1–17. DOI: 10.1093/gigascience/giad029. URL: https://inria.hal.science/hal-03769993.

[15] 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. URL: https://hal.inria.fr/hal-03025572.

[16] V. Guyomard, F. Fessant, T. Guyet, T. Bouadi and A. Termier. 'Generating robust counterfactual explanations'. In: ECML/PKDD - European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases. Turin (Italie), Italy, 2023, pp. 1–16. URL: https://hal.science/hal-04255500.

[17] V. Guyomard, F. Fessant, T. Guyet, T. Bouadi and A. Termier. 'Interactive Visualization of Counterfactual Explanations for Tabular Data'. In: ECML/PKDD 2023 - European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases. Vol. 14175. Lecture Notes in Computer Science. Turin, Italy: Springer Nature Switzerland, 17th Sept. 2023, pp. 330–334. DOI: 10.1007/978-3-031-43430-3_25. URL: https://hal.science/hal-04255496.

[18] J. Lehmann, P. Gattogi, D. Bhandiwad, S. Ferré and S. Vahdati. 'Language Models as Controlled Natural Language Semantic Parsers for Knowledge Graph Question Answering'. In: *Frontiers in Artificial Intelligence and Applications*. ECAI 2023 - 26th European Conference on Artificial Intelligence. Vol. 372. Frontiers in Artificial Intelligence and Applications. Krakow (Cracovie), Poland: IOS Press, 28th Sept. 2023, pp. 1348–1356. DOI: 10.3233/FAIA230411. URL: https://inria.hal.science/hal-04269089.

[19] G. G. Méndez, L. Galárraga, K. Chiluiza and P. Mendoza. 'Impressions and Strategies of Academic Advisors When Using a Grade Prediction Tool During Term Planning'. In: CHI 2023 - Conference on Human Factors in Computing Systems. Hamburg, Germany: ACM, 2023, pp. 1–18. DOI: 10.1145/3544548.3581575. URL: https://inria.hal.science/hal-04132566.

12.2 Publications of the year

International journals

[20] H. A. Ayats, P. Cellier and S. Ferré. 'Concepts of Neighbors and their Application to Instance-based Learning on Relational Data'. In: *International Journal of Approximate Reasoning* 164 (Oct. 2023), pp. 1–49. DOI: 10.1016/j.ijar.2023.109059. URL: https://inria.hal.science/hal-04246864.

[21] M. Durand, C. Largouët, L. Bonneau de Beaufort, J.-Y. Dourmad and C. Gaillard. 'A dataset to study group-housed sows' individual behaviours and production responses to different short-term events'. In: *Animal - Open Space* 2 (2023), p. 100039. DOI: 10.1016/j.anopes.2023.100039. URL: https://hal.inrae.fr/hal-04040702.

[22] M. Durand, C. Largouët, L. Bonneau de Beaufort, J.-Y. Dourmad and C. Gaillard. 'Estimation of gestating sows' welfare status based on machine learning methods and behavioral data'. In: *Scientific Reports* 13.1 (2023), p. 21042. DOI: 10.1038/s41598-023-46925-z. URL: https://hal.inrae.fr/hal-04320566.

[23] M. Durand, C. Largouët, L. Bonneau de Beaufort, J.-Y. Dourmad and C. Gaillard. 'Prediction of the daily nutrient requirements of gestating sows based on sensor data and machine-learning algorithms'. In: *Journal of Animal Science* 101 (2023), skad337. DOI: 10.1093/jas/skad337. URL: https://hal.inrae.fr/hal-04235874.

[24] O. Gauriau, L. Galárraga, F. Brun, A. Termier, L. Davadan and F. Joudelat. 'Comparing machine-learning models of different levels of complexity for crop protection: A look into the complexity-accuracy tradeoff'. In: *Smart Agricultural Technology* 7 (Mar. 2024), p. 100380. DOI: 10.1016/j.atech.2023.100380. URL: https://hal.science/hal-04382202.

[25] C. Gautrais, P. Cellier, T. Guyet, R. Quiniou and A. Termier. 'Sky-signatures: detecting and characterizing recurrent behavior in sequential data'. In: *Data Mining and Knowledge Discovery* (29th Aug. 2023). DOI: 10.1007/s10618-023-00949-1. URL: https://hal.science/hal-04401641.

[26] E. Germani, E. Fromont and C. Maumet. 'On the benefits of self-taught learning for brain decoding'. In: *GigaScience* 12 (3rd May 2023), pp. 1–17. DOI: 10.1093/gigascience/giad029. URL: https://inria.hal.science/hal-03769993.

International peer-reviewed conferences

[27] A. 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; IEEE, 18th Apr. 2023, pp. 1–6. DOI: 10.1109/ICTAI56018.2022.00090. URL: https://hal.science/hal-03890057.

[28] A. N. Amalou, E. Fromont and I. Puaut. 'CAWET: Context-Aware Worst-Case Execution Time Estimation Using Transformers'. In: *Leibniz International Proceedings in Informatics (LIPIcs), Volume 262, pp. 7:1-7:20, Schloss Dagstuhl - Leibniz-Zentrum für Informatik*. ECRTS 2023 - 35th Euromicro Conference on Real-Time Systems. Vol. 262. Vienne, Austria: Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 3rd July 2023, 7:1–7:20. DOI: 10.4230/LIPIcs.ECRTS.2023.7. URL: https://hal.science/hal-04148587.

[29] A. N. Amalou, E. Fromont and I. Puaut. 'Fast and Accurate Context-Aware Basic Block Timing Prediction using Transformers'. In: *Proceedings of the ACM SIGPLAN 2024 International Conference on Compiler Construction*. CC 2024 - ACM SIGPLAN 33rd International Conference on Compiler Construction. Edimbourg, United Kingdom: ACM, 2nd Mar. 2024, pp. 1–12. DOI: 10.1145/nnnnnn.nnnnnnn. URL: https://hal.science/hal-04406073.

[30] I. Calviac, O. Sankur and F. Schwarzentruber. 'Improved Complexity Results and an Efficient Solution for Connected Multi-Agent Path Finding'. In: *Proc. of the 22nd International Conference on AutonomousAgents and Multiagent Systems (AAMAS 2023)*. AAMAS 2023 - 22nd International Conference on Autonomous Agents and Multiagent Systems. London, United Kingdom, 2023, pp. 1–9. URL: https://hal.science/hal-04075393.

[31] S. Corbillé, E. Anquetil and E. Fromont. 'Precise Segmentation for Children Handwriting Analysis by Combining Multiple Deep Models with Online Knowledge'. In: ICDAR 2023 - 17th International Conference on Document Analysis and Recognition. San José, United States, 21st Aug. 2023, pp. 1–18. URL: https://hal.science/hal-04142592.

[32] S. Felton, É. Fromont and E. Marchand. 'Deep metric learning for visual servoing: when pose and image meet in latent space'. In: ICRA 2023 - IEEE International Conference on Robotics and Automation. London, United Kingdom: IEEE, 29th May 2023, pp. 741–747. DOI: 10.1109/ICRA48891.2023.10160963. URL: https://inria.hal.science/hal-04003126.

[33] S. Ferré. 'Dexteris: Data Exploration and Transformation with a Guided Query Builder Approach'. In: DEXA 2023 - 34th International Conference on Database and Expert Systems Applications. Vol. LNCS-14146. Penang, Malaysia, Malaysia: Springer, 2023, pp. 361–376. DOI: 10.1007/978-3-031-39847-6_29. URL: https://inria.hal.science/hal-04186117.

[34] S. Ferré. 'Graph-FCA Meets Pattern Structures'. In: *Lecture Notes in Artificial Intelligence*. ICFCA 2023 - 17th International Conference on Formal Concept Analysis. Vol. LNAI-13934. Kassel, Germany: Springer, 2023, pp. 33–48. DOI: 10.1007/978-3-031-35949-1_3. URL: https://inria.hal.science/hal-04186101.

[35] S. Ferré. 'Le principe MDL au service de l'automatisation de tâches uniques d'abstraction et de raisonnement (ARC) à partir de peu d'exemples'. In: *EGC 2023*. EGC 2023 - Extraction et Gestion des Connaissances. Vol. Revue des Nouvelles Technologies de l'Information, RNTI-E-39. Lyon, France, 2023, pp. 235–246. URL: https://inria.hal.science/hal-04186090.

[36] L. Galárraga. 'Effects of Locality and Rule Language on Explanations for Knowledge Graph Embeddings'. In: IDA 2023 - Advances in Intelligent Data Analysis XXI. Vol. 13876. Lecture Notes in Computer Science. Louvain-la-Neuve, Belgium: Springer Nature Switzerland, 1st Apr. 2023, pp. 143–155. DOI: 10.1007/978-3-031-30047-9_12. URL: https://inria.hal.science/hal-04132499.

[37] L. Galárraga, D. Hernández, A. Katim and K. Hose. 'Visualizing How-Provenance Explanations for SPARQL Queries'. In: WWW 2023 - ACM International World Wide Web Conference. Austin, United States: ACM, 2023, pp. 212–216. DOI: 10.1145/3543873.3587350. URL: https://inria.hal.science/hal-04386268.

[38] V. Guyomard, F. Fessant, T. Guyet, T. Bouadi and A. Termier. 'Generating robust counterfactual explanations'. In: ECML/PKDD - European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases. Turin (Italie), Italy, 2023, pp. 1–16. URL: https://hal.science/hal-04255500.

[39] V. Guyomard, F. Fessant, T. Guyet, T. Bouadi and A. Termier. 'Interactive Visualization of Counterfactual Explanations for Tabular Data'. In: ECML/PKDD 2023 - European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases. Vol. 14175. Lecture Notes in Computer Science. Turin, Italy: Springer Nature Switzerland, 17th Sept. 2023, pp. 330–334. DOI: 10.1007/978-3-031-43430-3_25. URL: https://hal.science/hal-04255496.

[40] G. Kelodjou, L. Rozé, V. Masson, L. Galárraga, R. Gaudel, M. Tchuente and A. Termier. 'Shaping Up SHAP: Enhancing Stability through Layer-Wise Neighbor Selection'. In: AAAI 2024 - 38th Annual AAAI Conference on Artificial Intelligence. Vancouver, Canada, 2024, pp. 1–10. URL: https://inria.hal.science/hal-04413022.

[41] J. Lehmann, P. Gattogi, D. Bhandiwad, S. Ferré and S. Vahdati. 'Language Models as Controlled Natural Language Semantic Parsers for Knowledge Graph Question Answering'. In: *Frontiers in Artificial Intelligence and Applications*. ECAI 2023 - 26th European Conference on Artificial Intelligence. Vol. 372. Frontiers in Artificial Intelligence and Applications. Krakow (Cracovie), Poland: IOS Press, 28th Sept. 2023, pp. 1348–1356. DOI: 10.3233/FAIA230411. URL: https://inria.hal.science/hal-04269089.

[42] G. G. Méndez, L. Galárraga, K. Chiluiza and P. Mendoza. 'Impressions and Strategies of Academic Advisors When Using a Grade Prediction Tool During Term Planning'. In: CHI 2023 - Conference on Human Factors in Computing Systems. Hamburg, Germany: ACM, 2023, pp. 1–18. DOI: 10.1145/3544548.3581575. URL: https://inria.hal.science/hal-04132566.

[43] O. Pelgrin, R. Taelman, L. Galárraga and K. Hose. 'GLENDA: Querying over RDF Archives with SPARQL'. In: ESWC 2023 - 20th International Conference on The Semantic Web. Vol. 13998. Lecture Notes in Computer Science. Hersonisos, Crete, Greece: Springer Nature Switzerland, 21st Oct. 2023, pp. 75–80. DOI: 10.1007/978-3-031-43458-7_14. URL: https://inria.hal.science/hal-04388974.

[44] O. Pelgrin, R. Taelman, L. Galárraga and K. Hose. 'Scaling Large RDF Archives To Very Long Histories'. In: ICSC 2023 - IEEE 17th International Conference on Semantic Computing. Laguna Hills, United States: IEEE, 2023, pp. 41–48. DOI: 10.1109/ICSC56153.2023.00013. URL: https://inria.hal.science/hal-04388912.

National peer-reviewed Conferences

[45] A. Chaffin and J. Delaunay. '"Honey, Tell Me What's Wrong", Explicabilité Globale des Modèles de TAL par la Génération Coopérative'. In: *Actes de CORIA-TALN 2023. Actes de la 30e Conférence sur le Traitement Automatique des Langues Naturelles (TALN),*
volume 1 : travaux de recherche originaux – articles longs. CORIA TALN RJCRI RECITAL 2023 - 18e Conférence en Recherche d'Information et Applications
16e Rencontres Jeunes Chercheurs en RI
30e Conférence sur le Traitement Automatique des Langues Naturelles
25e Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues. Paris, France: ATALA, 2023, pp. 105–122. URL: https://hal.science/hal-04130137.

[46] V. Guyomard, F. Wallyn, F. Fessant, T. Guyet, T. Bouadi and A. Termier. 'Visualiser des explications contrefactuelles pour des données tabulaires'. In: *Revue des Nouvelles Technologies de l'Information*. EGC 2023 - Conférence Extraction et Gestion des Connaissances. Vol. RNTI-E-39. Lyon, France, 16th Jan. 2023, pp. 557–564. URL: https://hal.science/hal-04181257.

Conferences without proceedings

[47] L. Cornanguer, C. Largouët, L. Rozé and A. Termier. 'Persistence-Based Discretization for Learning Discrete Event Systems from Time Series'. In: MLmDS 2023 - AAAI Workshop When Machine Learning meets Dynamical Systems: Theory and Applications. Washington (DC), United States, 13th Feb. 2023, pp. 1–6. URL: https://inria.hal.science/hal-03934438.

[48] J. Delaunay, C. Largouët, L. Galárraga and N. V. Berkel. 'Adaptation of AI Explanations to Users' Roles'. In: HCXAI 2023 - Workshop on Human-Centered Explainable AI. Hamburg, Germany, 2023, pp. 1–7. URL: https://inria.hal.science/hal-04388942.

[49] O. Pelgrin, R. Taelman, L. Galárraga and K. Hose. 'The Need for Better RDF Archiving Benchmarks'. In: MEPDaW 2023 - Managing the Evolution and Preservation of the Data Web. Athens, Greece, 2023, pp. 1–5. URL: https://inria.hal.science/hal-04389014.

Scientific book chapters

[50] P. Cellier, M. Ducassé, S. Ferré, O. Ridoux and W. E. Wong. 'Data Mining-Based Techniques for Software Fault Localization'. In: *Handbook of Software Fault Localization*. 1. Wiley, 20th Apr. 2023, Chapitre 7. DOI: 10.1002/9781119880929.ch7. URL: https://hal.science/hal-04403506.

Doctoral dissertations and habilitation theses

[51] A. N. Amalou. 'Machine learning for timing estimation'. Université de Rennes, 12th Dec. 2023. URL: https://hal.science/tel-04406029.

[52] S. Corbillé. 'Integrating Explicit Knowledge with Deep Learning for Children's Handwriting Recognition and Segmentation'. Université de Rennes, 28th June 2023. URL: https://theses.hal.science/tel-04236705.

[53] L. Cornanguer. 'Timed automata learning from time series'. Université de Rennes, 2nd Nov. 2023. URL: https://theses.hal.science/tel-04390077.

[54] A. Voyez and A. Voyez. 'Privacy risk analysis of large-scale temporal data : application to electricity consumption data'. Université de Rennes, 11th July 2023. URL: https://theses.hal.science/tel-04215588.

Reports & preprints

[55] E. Germani, E. Fromont and C. Maumet. *Uncovering communities of pipelines in the task-fMRI analytical space*. 8th Dec. 2023. URL: https://hal.science/hal-04331232.

[56] E. Germani, E. Fromont, P. Maurel and C. Maumet. *The HCP multi-pipeline dataset: an opportunity to investigate analytical variability in fMRI data analysis*. 21st Dec. 2023. URL: https://inserm.hal.science/inserm-04356768.

Other scientific publications

[57] L. Brossard, C. Largouët and L. Bonneau de Beaufort. 'Real-time combination of observed growth and feed intake performance with performance simulated by InraPorc® to apply precision feeding to growing pigs'. In: 55. Journées de la Recherche Porcine (JRP). Vol. 14. 55èmes Journées de la recherche porcine 5. Saint-Malo, France: Ifip, 2023, 133-134 | 670–671. DOI: 10.1016/j.anscip.2023.06.037. URL: https://hal.inrae.fr/hal-04094219.

[58] P. Cellier. 'Scikit-mine: A pattern mining library in Python'. In: SMiLe 2023 - Spring workshop on Mining and Learning. Sint-Michielgestel, Netherlands, 2023, pp. 1–1. URL: https://hal.science/hal-04405499.

[59] M. Durand, C. Largouët, L. Bonneau de Beaufort, J.-Y. Dourmad and C. Gaillard. 'Prediction of daily nutritional requirements of gestating sows based on their behaviour and machine learning methods'. In: ESPHM - 14th European symposium of porcine health management. Thessalokini, Greece, 2023, pp. 1–1. URL: https://hal.inrae.fr/hal-04119697.

[60] E. Germani, E. Fromont and C. Maumet. 'Exploring variability patterns in the task-fMRI analytical space'. In: OHBM 2023 - 29th Annual Meeting of the Organization for Human Brain Mapping. Montreal, Canada, July 2023. URL: https://inria.hal.science/hal-03991042.

[61] E. Germani, E. Fromont and C. Maumet. 'Representation learning for more reproducible fMRI data analyses'. In: IABM 2023 - Colloque Français d'Intelligence Artificielle en Imagerie Biomédical. Paris, France, 30th Mar. 2023, p. 1. URL: https://inria.hal.science/hal-04068797.