

Pierre-Edouard PORTIER, PhD

**Home Address:**

1 rue Marteret
69100 VILLEURBANNE
France

Email:

pierre-edouard.portier@insa-lyon.fr ;
pierre.edouard.portier@gmail.com

Homepage:

<https://peportier.me>

LinkedIn:

Cell: +33 (0)6 77 17 54 54

DoB:

30th November 1984

<https://www.linkedin.com/in/peportier/>

One-page summary

Since 2011, Pierre-Edouard Portier is *associate professor at INSA Lyon*, France's leading post-baccalaureate engineering school, within the Computer Science & Information Technologies (CS&IT) department where he also studied and obtained an engineering degree. After his MSc, he obtained in 2010 a PhD degree in computer science from INSA Lyon.

Research topics: His research addresses topics in the fields of *machine learning* and soft computing techniques applied to big data analytics (e.g., anomaly detection, XAI for crash prediction, traffic forecasting...), *natural language processing* (e.g., relation extraction, knowledge graph completion...), *semantic web search* and *document engineering*. His research interest lies in the discovery, modeling, and representation of knowledge to be integrated into a data analytics process.

Management & community activities: He is currently *deputy director in charge of corporate relations for INSA Lyon CS&IT*. He is an *active member of IRIXYS*, an International Research and Innovation Center in Intelligent Digital Systems. He has also an activity in *academic consulting* as lead scientist in the context of collaborations with several companies. He has been a member of various program committees and reviewers for journals, conferences, and workshops such as Elsevier's Expert Systems with Applications, Wiley's Expert Systems, Springer's EPJ Data Science, the European Conference on Information Retrieval, ACM I-KNOW conference, IEEE International Conference on Web Intelligence, IEEE Transactions on Neural Networks and Learning Systems, etc.

From September 2018 to September 2020, he has been *deputy director in charge of digital strategy at INSA Lyon*. Until September 2018, he managed the organization of seminars with industrial partners invited to share their expertise with INSA Lyon CS&IT students. From 2015 to 2019, he was responsible for the 4th year of studies at INSA Lyon CS&IT. During the academic year 2015-2016, he co-managed its evolution towards a more pronounced project-based learning pedagogy. From 2012 to 2018 and since 2022, he is an elected member of INSA Lyon CS&IT board. From 2012 to 2019, he was an elected representative for the LIRIS laboratory board.

Lecturing: Pierre-Edouard Portier is teaching artificial intelligence, data analytics, agile software development methodologies and formal methods for software quality at INSA Lyon CS&IT. He is supervising or has supervised 9 PhD students in computer science at LIRIS laboratory.

Key publications:

- Pierre-Yves Genest, Pierre-Edouard Portier, Elöd Egyed-Zsigmond, Martino Lovisetto: **Linked-DocRED – Enhancing DocRED with Entity-Linking to Evaluate End-To-End Document-Level Information Extraction**. ACM SIGIR 2023.
- Thomas Veran, Pierre-Edouard Portier, François Fouquet: **Interpretable hierarchical symbolic regression for safety-critical systems with an application to highway crash prediction**. Engineering Applications of Artificial Intelligence 2023.
- Pierre-Yves Genest, Pierre-Edouard Portier, Elöd Egyed-Zsigmond, Laurent-Walter Goix: **PromptORE – A Novel Approach Towards Fully Unsupervised Relation Extraction**. CIKM 2022.
- Yvan Lucas, Pierre-Edouard Portier, Lea Laporte, Sylvie Calabretto, Olivier Caelen, Liyun He-Guelton, Michael Granitzer: **Towards automated feature engineering for credit card fraud detection using multi-perspective HMMs**. Future Generation Comp. Syst. 2020.
- Johannes Jurgovsky, Michael Granitzer, Konstantin Ziegler, Sylvie Calabretto, Pierre-Edouard Portier, et al.: **Sequence Classification for Credit-Card Fraud Detection**. Expert Systems with Applications, Elsevier, 2018
- Vincent Barrellon, Pierre-Edouard Portier, Sylvie Calabretto, Olivier Ferret: **Linear Extended Annotation Graphs**. ACM Symposium on Document Engineering 2017
- Vincent Barrellon, Pierre-Edouard Portier, Sylvie Calabretto, Olivier Ferret: **Schema-aware Extended Annotation Graphs**. ACM Symposium on Document Engineering 2016
- Mazen Alsarem, Pierre-Edouard Portier, Sylvie Calabretto, Harald Kosch: **Ranking Entities in the Age of Two Webs, an Application to Semantic Snippets**. ESWC 2015: 541-555
- Pierre-Edouard Portier, Sylvie Calabretto: **DINAH, A Philological Platform for the Construction of Multi-structured Documents**. ECDL 2010: 364-375

Patent:

Caelen, O., He-Guelton, L., Portier, P.E., Granitzer, M., Ziegler, K. and Jurgovsky, J., Worldline SA, 2020. **Machine learning system for various computer applications**. U.S. Patent Application 16/632,267.

2012-09-01 / 2016-09-01 Semantic Web Search: Pierre-Edouard Portier supervised Mazen Alsarem's PhD thesis, which aimed to enhance traditional web search engine results with information from the web of data through a concept called "semantic snippet". To address the challenge of inaccurate or irrelevant discovered entities in web pages, Portier, Alsarem, and Calabretto developed the LDRANK algorithm [1,2,3]. LDRANK extended the PageRank algorithm and ranked entities based on their proximity to the user's query, resulting in a robust entity ranking system. The algorithm's efficiency was demonstrated through a dataset collected via crowdsourcing. The thesis was supported by a grant from the French Region Rhône-Alpes.

2014-09-01 / 2017-09-01 Multilingual Semantic Web Search: Pierre-Edouard Portier extended the work done on LDRANK to higher-order tensors as part of the supervision of Melkamu Beyene's PhD thesis [4,5]. This adaptation allowed the algorithm to be used in a multilingual context. Beyene's thesis was funded by a collaboration between France and Ethiopia, with Portier serving as the main supervisor.

2013-09-01 / 2017-09-01 Extended Annotation Graphs: Pierre-Edouard Portier collaborated with Prof. Sylvie Calabretto and Prof. Olivier Ferret in co-supervising Vincent Barrellon's PhD thesis, which focused on digital scholarly edition. Together, they developed the annotation formalism eAG (Extended Annotation Graphs) and its schema language, SeAG [6,7], in collaboration with humanities researchers working on a digital scholarly edition of the Encyclopédie of Diderot and D'Alembert. The scholars needed to annotate documents using an evolving schema, and XML documents with their inherent tree structure did not meet their needs. To address this, Barrellon, Portier, Calabretto, and Ferret proposed an efficient algorithm for validating acyclic graphs based on an expressive schema. They also designed a bi-directional algebra to propagate the effects of a schema modification on existing annotations semi-automatically. Finally, they defined a markup syntax, like XML, to represent eAG documents. The thesis was supported by a grant from the French Region Rhône-Alpes.

2015-09-01 / 2019-09-01 Hidden Markov Model Feature Engineering for Fraud Detection: Pierre-Edouard Portier collaborated with Prof. Sylvie Calabretto, Prof. Harald Kosch, and Dr. Léa Laporte in co-supervising Yvan Lucas' PhD Thesis, which aimed to improve credit card fraud detection. As a member of the international research center IRIXYS, Pierre-Edouard was part of a partnership with one of Europe's major companies monitoring credit card transactions. Two PhD theses, including Yvan Lucas', were funded based on this partnership and focused on credit card fraud detection. Yvan Lucas' thesis introduced an innovative way to model the behavior of both fraudsters and normal cardholders. The new algorithm integrated representations of sequential behaviors learned by probabilistic generative models (e.g., hidden Markov model) into a flexible machine learning model (e.g., random forest). The developed approach outperformed the state of the art due to its ability to model both genuine and fraudulent behaviors [10,11,12].

2015-09-01 / 2019-09-01 LSTM for Credit Card Fraud Detection: Pierre-Edouard Portier collaborated with Prof. Michael Granitzer and Prof. Sylvie Calabretto in co-supervising Johannes Jurgovsky's PhD thesis, which aimed to improve credit card fraud detection using neural network-based approaches. This thesis was also funded by the same project that funded Yvan Lucas' thesis. The work had two main contributions. First, it introduced a method to integrate embeddings of knowledge graphs into a machine learning model to improve credit card fraud detection by providing a richer representation of geographical information linked to cardholders and payment terminals [9]. Second, it produced a well-cited reference study on the use of LSTM neural networks to represent cardholders' behaviors and improve fraud detection [8]. The work resulted in the filing of a patent [15].

2019-09-01 / 2022-09-01 Interpretable Hierarchical Symbolic Regression: Pierre-Edouard Portier collaborated with Professor Jean-Marc Petit and Dr. François Fouquet to co-supervise Thomas Véran's PhD thesis. The thesis aimed to develop a data-centric approach to discover the risk factors associated with highway accidents. This project was funded in collaboration with an industrial partner: Data New Road. Data New Road is a joint venture between IT company Amiltone and APRR, a subsidiary of Eiffage, one of Europe's largest motorway managers. The team developed a Bayesian hierarchical model to identify risk factors related to highway accidents [13]. This work is a contribution to the field of eXplainable Machine Learning (XAI), and its originality lies in two key areas. First, they used the Shapley values of a black box model (e.g., XGBM) to automatically discover a hierarchical structure, leveraging the features' importance. Second, they used an auto-adaptive polynomial neural network to automatically detect significant feature interactions. To achieve an even better balance between predictive performance and interpretability, the team introduced an innovative symbolic regression model [16]. This model rests on a dynamic interpretive process that favors a progressive analysis of numerous instructive models ranked by increasing complexity.

2021-09-01 / ... Natural Language Processing for Open Relation Extraction and Knowledge Graph Completion: Pierre-Edouard Portier is currently co-supervising Pierre-Yves Genest's PhD thesis alongside Elöd Egyed-Zsigmond. This project is funded through a collaboration with Alteca, a French IT company. The main goal of this thesis is to make a significant contribution to the field of natural language processing (NLP) by improving knowledge graph completion from plain text documents. A first step towards this objective has been the development of an unsupervised relation extraction model called PromptORE [14], short for "Prompt-based Open Relation Extraction." PromptORE is a highly innovative model that adapts the prompt-tuning paradigm to work in an unsupervised setting, enabling the creation of embeddings for sentences expressing a relation. Notably, PromptORE is the first unsupervised relation extraction model that does not require any hyperparameter tuning and consistently outperforms state-of-the-art models. Next, they addressed the limitations of existing datasets for Information Extraction (IE) pipeline evaluation by introducing Linked-DocRED [17]. This dataset is the first manually annotated, large-scale, document-level IE dataset featuring annotations for entities, coreferences, relations, and entity-linking. They achieve this by enhancing the widely used DocRED dataset with entity-linking labels generated through a semi-automatic process, ensuring high-quality annotations. Alongside this, they propose a complete framework of metrics and an entity-centric metric for entity-linking evaluation. This work significantly contributes to the field by providing a robust benchmark for end-to-end IE pipeline evaluation, enabling further research and development in information extraction pipelines.

2019-09-01 / ... DataValor Academic Consulting: Pierre-Edouard Portier is a member of the DataValor team, contributing to the acceleration of the valorization of the LIRIS laboratory research results concerning data analysis. DataValor is affiliated with INSAVALOR, a subsidiary of INSA Lyon, and its primary objective is to offer expert services to companies for data valorization. Within this framework, Pierre-Edouard has been the scientific director of a project with the joint venture Data New Road, which aimed to develop predictive models of highway traffic. Additionally, he initiated and co-supervised Thomas Véran's CIFRE thesis with Prof. Jean-Marc Petit, focusing on the predictive analysis of accident causes on highways.

Furthermore, Pierre-Edouard's involvement in DataValor has enabled him to conduct scientific consulting and training assignments for companies that face data valorization problems. For instance, he collaborated with a large French bank in creating a data science training program and helped them tackle their machine learning and data science obstacles through several consulting missions. He also supported them in recruiting data scientists.

He is also the scientific lead for a project with UGECAM (an organization affiliated with French Health Insurance) focused on analyzing pediatric sleep respiratory polygraphy. Polysomnography is the standard method for evaluating pediatric sleep, but it is complex and limited to major hospital centers. Respiratory polygraphy (PG) offers a simpler alternative for diagnosing sleep apnea, but current automatic analyses often fail in children. The team has studied the errors in commercial automatic analysis models on pediatric recordings and concluded that their reliability is insufficient. This work led to a poster publication at the Sleep Congress in 2022 [18]. The project now aims to develop an automatic PG analysis model for pediatrics based on the expertise of practitioners. The method will involve observing expert analysis methods, formalizing them into algorithms, and facilitating data interpretation. An initial step involves identifying children's wake phases from PG data, in the absence of electroencephalogram (EEG) information. This project has received CNRS DECLIC funding, a pre-maturation program designed to support the early stages of this innovative project, with the aim of reaching sufficient maturity for transfer to industry or the creation of a company.

Teaching

Advanced Algorithm for AI: A module for M1 students at INSA Lyon CS&IT department. It first introduces algorithms for solving combinatorial problems formulated as shortest path searches (A^* , IDA*, etc.) or configuration searches on a state graph (simulated annealing, tabu search, etc.). It also presents the PageRank algorithm for ordering the nodes of a graph according to its structure. It also introduces Gödel's incompleteness theorems and their link with Turing's work. Finally, it presents the basics of combinatorial game theory.

Formal approach for building software: A module for M2 students at INSA Lyon CS&IT department, in which I introduce how to write programs correct by construction by transforming a formal specification into a program. We follow the approach promoted by Dijkstra and based on weakest precondition predicates.

Object oriented conception and Agile development: A module for M2 students at INSA Lyon CS&IT department, in which we have students develop in groups of 6 an application for building optimal city-wide delivery rounds.

Introduction to Machine Learning: A module for M2 students at INSA Lyon CS&IT department. I designed it and implemented it for the first time in the academic year 2021-2022. The discipline of machine learning develops algorithms and methods to discover regularities in multidimensional data to, among other things, automate prediction. This module introduces the subfield of supervised learning, which consists of predicting the association between a label and an object described by several variables. It presents the fundamental concepts of supervised learning (e.g., regularization, bias, variance...) and shows their various interconnections in the context of algorithmic developments that allow the analysis of datasets with a primarily predictive aim. Thus, theoretical propositions lead to the writing of programs that implement or use some essential models of supervised learning (e.g., regularized linear regression, kernel ridge regression, Nyström approximation...). The course is accompanied by manipulatable notebooks written in the R programming language. These systematic exercises link fundamental concepts to their applications in data analysis projects.

Indexing large volume of textual data: A module for M2 students at INSA Lyon CS&IT department, where I introduce the main algorithms for indexing and querying large amount of textual data.

Collective and general interest responsibilities

2020-09-01 / ... Deputy Director in charge of Corporate Relation at INSA Lyon CS&IT department: Pierre-Edouard Portier, as Deputy Director responsible for Corporate Relations at INSA Lyon's Computer Science and Information Technology department, has established and is leading a club of industrial partners. This club is dedicated to building and strengthening relationships with a select group of companies recognized for their excellence and values, such as Avanade, BioMérieux, CGI, Crédit Agricole, Deloitte, L'Oréal, Société Générale, SAP, Sopra Steria, Volvo, Onepoint, among others.

Under his guidance, Pierre-Edouard and his team organize a variety of events, including conferences, afterworks, and seminars, providing opportunities for companies, students, and faculty to interact. The interactions that take place within this partner's club are of high quality and serve as a model for the INSA group's partnership strategy development. It has become a central element for the department.

2016-09-01 / ... IRIXYS International Research & Innovation Center on Digital Intelligent Systems: Pierre-Edouard Portier has been an active member of IRIXYS, an International Research & Innovation Center on Digital Intelligent Systems, since 2016. The center, directed by Professors Ernesto Damiani (Università degli Studi di Milano), Lionel Brunie (INSA Lyon), and Harald Kosch (University of Passau), is dedicated to advancing research in the field of digital intelligent systems. Pierre-Edouard Portier has played a central role in supervising the PhD theses of Mazen Alsarem, Yvan Lucas, Johannes Jurgosky, and Pierre-Yves Genest. Additionally, Pierre-Edouard Portier is an active participant in the bi-annual workshops organized by IRIXYS. These workshops aim to steer various Franco-Italian-German projects and guide the evolution of PhD thesis work of numerous co-tutored students. One of Pierre-Edouard Portier's notable achievements was his involvement in a multi-year contract with a large European company specializing in the monitoring of banking transactions. He played a key role in supervising the PhD thesis of Yvan Lucas and Johannes Jurgovsky in this context.

2020-09-01 / ... Deputy Director CISR, Network Services Inter-Universities Center, in Lyon, France: As of September 2020, Pierre-Edouard Portier has been serving as the Deputy Director of the Network Services Inter-Universities Center (CISR), where he represents the interests of INSA Lyon. This center serves as a central hub for all university projects related to network and infrastructures for data storage and computation in the Lyon region of France. Pierre-Edouard Portier is responsible for overseeing the management of activities related to the maintenance and evolution of the INSA Lyon computer network. As the Deputy Director of this center, he plays a key role in setting up a major project to construct a data center on the LyonTech La Doua campus, which will house all the site's digital research infrastructures.

2018-09-01 / 2020-09-01 Deputy Director in charge of Digital at INSA Lyon: Pierre-Edouard Portier, as Deputy Director in charge of digital, has led various projects to transform the digital landscape at INSA Lyon. These include the integration of an electronic document management and collaborative work platform, positioning of INSA Lyon as a pilot site for a new national school management software, and support for start-ups that arise from student initiatives. Pierre-Edouard Portier's contributions to the institution's digital transformation are noteworthy. He was responsible for the evolution of the digital master plan and played an active role in steering the Information Systems Department's evolution at INSA. He also helped manage the COVID crisis to ensure that services continued uninterrupted and assisted students and staff who were facing difficulties due to limited access to digital resources.

2012-2018; 2022-... — Elected member of INSA Lyon CS&IT department board

2012-2019 — Elected member of LIRIS laboratory board

Supervised PhD Thesis

- **Dr Mazen ALSAREM** (2012-2016); supervision rate: 80%; co-supervisor: Pr Sylvie CALABRETTO
- **Dr Melkamu BEYENE** (2014-2017); supervision rate: 80%; co-supervisor: Pr Sylvie CALABRETTO
- **Dr Vincent BARRELLON** (2013-2017); supervision rate: 80%; co-supervisors: Pr Sylvie CALABRETTO, Pr Olivier FERRET
- **Dr Yvan LUCAS** (2015-2019); supervision rate: 50%; co-supervisors: Pr Harald KOSCH, Pr Sylvie CALABRETTO, Dr Léa LAPORTE
- **Dr Johannes JURGOVSKY** (2015-2019); supervision rate: 50%; co-supervisors: Pr Michael GRANITZER, Pr Sylvie CALABRETTO
- **Dr Thomas VÉРАН** (2019-2022); supervision rate: 95%; co-supervisor: Pr Jean-Marc PETIT
- **Pierre-Yves GENEST** (2021-); supervision rate: 50%; co-supervisor: Dr Elöd EGYED-ZSIGMOND
- **Johan LEYDET** (2022-); supervision rate: 33%; co-supervisors: Dr Elöd EGYED-ZSIGMOND, Dr Diana NURBAKOVA
- **Anais BOUMENDIL** (2022-); supervision rate: 25%; co-supervisors: Pr Frédéric LE MOUËL, Dr Walid BECHKIT, Dr Malcolm EGAN

References:

- [1] Alsarem, M., Portier, P.E., Calabretto, S. and Kosch, H., 2015, March. Ordonnancement d'entités appliqué à la construction de snippets sémantiques. In *CORIA 2015: Conférence en Recherche d'Information et Applications* (pp. 253-268).
- [2] Alsarem, M., Portier, P.E., Calabretto, S. and Kosch, H., 2015. Ordonnancement d'entités pour la rencontre du web des documents et du web des données. *Document numérique*, 18(2), pp.123-154.
- [3] Alsarem, M., Portier, P.E., Calabretto, S. and Kosch, H., 2015, May. Ranking entities in the age of two webs, an application to semantic snippets. In *European Semantic Web Conference* (pp. 541-555). Springer, Cham.
- [4] Beyene, M., Portier, P.E., Atnafu, S. and Calabretto, S., 2016, November. Dataset linking in a multilingual linked open data context. In *Proceedings of the 8th International Conference on Management of Digital EcoSystems* (pp. 149-157).
- [5] Beyene, M., Portier, P.E., Atnafu, S. and Calabretto, S., 2015, October. Tensor factorization for cross lingual entity co-reference resolution in the linked open data. In *Proceedings of the 7th International Conference on Management of computational and collective intelligence in Digital EcoSystems* (pp. 17-23).
- [6] Barrellon, V., Portier, P.E., Calabretto, S. and Ferret, O., 2016, September. Schema-aware extended annotation graphs. In *Proceedings of the 2016 ACM Symposium on Document Engineering* (pp. 45-54).
- [7] Barrellon, V., Portier, P.E., Calabretto, S. and Ferret, O., 2017, August. Linear extended annotation graphs. In *Proceedings of the 2017 ACM Symposium on Document Engineering* (pp. 9-18).
- [8] Jurgovsky, J., Granitzer, M., Ziegler, K., Calabretto, S., Portier, P.E., He-Guelton, L. and Caelen, O., 2018. Sequence classification for credit-card fraud detection. *Expert Systems with Applications*, 100, pp.234-245.
- [9] Ziegler, K., Caelen, O., Garchery, M., Granitzer, M., He-Guelton, L., Jurgovsky, J., Portier, P.E. and Zwicklbauer, S., 2017, June. Injecting semantic background knowledge into neural networks using graph embeddings. In *2017 IEEE 26th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE)* (pp. 200-205). IEEE.
- [10] Lucas, Y., Portier, P.E., Laporte, L., Calabretto, S., Caelen, O., He-Guelton, L. and Granitzer, M., 2019, April. Multiple perspectives HMM-based feature engineering for credit card fraud detection. In *Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing* (pp. 1359-1361).
- [11] Lucas, Y., Portier, P.E., Laporte, L., Calabretto, S., He-Guelton, L., Oblé, F. and Granitzer, M., 2019, June. Dataset shift quantification for credit card fraud detection. In *2019 IEEE second international conference on artificial intelligence and knowledge engineering (AIKE)* (pp. 97-100). IEEE.
- [12] Lucas, Y., Portier, P.E., Laporte, L., He-Guelton, L., Caelen, O., Granitzer, M. and Calabretto, S., 2020. Towards automated feature engineering for credit card fraud detection using multi-perspective HMMs. *Future Generation Computer Systems*, 102, pp.393-402.
- [13] Veran, T., Portier, P.E. and Fouquet, F., 2020, December. Crash prediction for a French highway network with an XAI-informed Bayesian hierarchical model. In *2020 IEEE International Conference on Big Data (Big Data)* (pp. 1256-1265).
- [14] Genest, P.Y., Portier, P.E., Egyed-Zsigmond, E and Goix, L.W., 2022, October. PromptORE — A novel approach towards fully unsupervised relation extraction. *CIKM 2022*.
- [15] Caelen, O., He-Guelton, L., Portier, P.E., Granitzer, M., Ziegler, K. and Jurgovsky, J., Worldline SA, 2020. Machine learning system for various computer applications. U.S. Patent Application 16/632,267.
- [16] Veran, T., Portier, P.E., Fouquet, F., 2023. Interpretable hierarchical symbolic regression for safety-critical systems with an application to highway crash prediction. *Engineering Applications of Artificial Intelligence* 2023.
- [17] Genest, P.Y., Portier, P.E., Egyed-Zsigmond, E and Lovisetto, M, 2023, July. Linked-DocRED – Enhancing DocRED with Entity-Linking to Evaluate End-To-End Document-Level Information Extraction. *ACM SIGIR 2023*.
- [18] Barrellon, V., Bertin, B., Portier, P. E., Depagne, C., De La Gastine, G., & Stagnara, A. (2023). Quelle fiabilité dans l'analyse automatique des polygraphies du sommeil (PG) en pédiatrie ? *Médecine du Sommeil*, 20(1), 69.