# Youcef REMIL, Ph.D

Detailed Curriculum Vitae

⊠ youcef.remil.cs@gmail.com DinkedIn Account

## Summary

I earned my engineering and Master of Research degrees from the Higher National School of Computer Science (ESI ex.INI), a leading institution for computer science education in Algeria. More recently, I completed my CIFRE Ph.D. thesis at INSA Lyon, which was conducted in collaboration with Infologic company and financially supported by them. My doctoral research primarily focused on the development of explainable AIOps (Artificial Intelligence for Operating Systems) software maintenance solutions. In a broader sense, I have a deep-seated interest in the fields of Machine Learning and Data Mining, particularly as they apply to Software Engineering (AI for SE). I possess a strong skill set in comprehending, processing, analyzing, and extracting valuable insights from diverse and intricate datasets. I harbor an inherent passion for tackling complex problems and take pleasure in delving into optimization algorithms, challenges in data mining, and mathematical problem-solving. My contributions within my thesis encompassed various aspects, including the overall structuring of the AIOps research domain, the application of pattern mining algorithms for SQL workload analysis, the identification of Java memory-related issues, and the utilization of machine learning techniques for bug triage and efficient similarity search in crash deduplication. Presently, I am employed as a research and data scientist at Infologic, where I focus on applied research in the realm of software maintenance.

# Work Experience

#### Nov. Research and Data Scientist, Infologic, Lyon, France.

2020-Present Infologic is one of France's leading providers of ERP solutions for the agri-food, health nutrition, and cosmetic sectors and has established a remarkable reputation over its 40 years in the market with a plethora of provided services and modules. Infologic is committed to providing comprehensive maintenance support of its ERP system across its entire server fleet, while ensuring the uninterrupted and reliable provision of services and mitigating any potential system faults and security threats. My doctoral thesis falls within this context.

Advisors: Mehdi KAYTOUE (HDR), Anes BENDIMERAD (Ph.D)

#### **Detailed achievements:**

- Conducting my CIFRE thesis project to study, compare and organize cutting-edge solutions within the emerging field of AlOps.
- Publishing research work in prestigious conferences and journals in Machine Learning and Software Engineering fields: ASE (Core A\*), ESEC-FSE (Core A\*), ICSE (Core A\*), TOSEM (Core A\*), DSAA (Core A), ICDMW (AIOps Workshop), EGC (National Conf).
- Implementing robust feature engineering tools to process and normalize rough and unstructured data (e.g., SQL queries parser, Stack traces parser, etc.).
- Designing a scalable and explainable model for incident assignment model to streamline the routing of maintenance tickets to the most suitable teams and individuals.
- Innovating a fast and efficient retrieval system that combines locality-sensitive hashing and contrastif learning to identify similar bugs/incidents in vast historical datasets.
- Crafting tools for the diagnosis of common software maintenance issues, notably Java memory errors and SQL query performance problems, using pattern mining methods.
- Employing Natural Language Processing (NLP) techniques to dissect maintenance tickets through a variety of innovative approaches such as Topic and Language Modeling.
- Providing mentorship to research interns, guiding them in areas such as anomaly detection for time series and fine-tuning large language models for maintenance application.
- Collaborating in the establishment of an AIOps and MLOps infrastructure responsible for monitoring thousands of computer servers in IT environments.

#### Oct.-Nov. Teaching Assistant, Institut National des Sciences Appliquées INSA, Lyon, France.

- 2022 Intervening in the OT2 course/project (Data Mining and Machine Learning) for 5th-year computer science students, with a total duration of 16 hours.
  - Study of the crash deduplication problem in software engineering.
  - Conducting a theoretical and experimental comparison of various similarity measures in this context.
  - Exploring and Implementing the MinHash technique.
- Jan.-July **Research Intern**, *Laboratoire LIRIS*, Lyon, France.
  - 2020 Conducting my graduation project which involved an exploration and implementation of eXplainable Artificial Intelligence (XAI) methodologies.

Advisors: Céline ROBARDET (Pr), Marc PLANTEVIT (HDR), Anes BENDIMERAD (Ph.D) Detailed achievements:

- Conducting an in-depth analysis of the state-of-the-art to categorize XAI) techniques.
- Proposing a novel approach that integrates Exceptional Model Mining with the recognized LIME method to deliver interpretable summaries of contextualized subgroups of predictions.
- Implementing and testing the designed approach against relevant baselines.
- Publishing the approach at DSAA 2021 conference (Core A).

#### Sep.-Oct. Network Engineering Intern, ALGERIE TELECOM, Mascara, Algérie.

- 2018 Collaborating with network engineers at the Transmission Center in Mascara, which operates under the purview of Algerie Telecom, the primary provider of fixed telephony, ADSL, and fiber optic services. **Detailed achievements:** 
  - Conducting a study on Next-Generation Networks (NGN).
  - Facilitating the migration to RMS Networks utilizing IP/MPLS technology.

# Educational Path

- 2020–2023 Ph.D. Computer Science, Institut National des Sciences Appliquées INSA, Lyon, France. Thesis: A Data Mining Perspective on Explainable AlOps with Applications to Software Maintenance Advisors: Jean-François BOULICAUT (Pr)
- 2019–2020 Master of Research. Computer Science, Ecole Nationale Superieure d'Informatique ESI, Algiers, Algeria.
   Thesis: Comprehensive Review of Methods for Explaining Black Box Models
   Grade: Excellent (18/20)
   Advisors: Karima BENATCHBA (Pr)
- 2015–2020 Engineering Degree. Computer Science , Ecole Nationale Superieure d'Informatique ESI, Algiers, Algeria.
   Thesis: Interpretable Summaries of Black Box predictions using Subgroup Discovery Grade: With honors (17.63/20) Advisors: Karima BENATCHBA (Pr)
- 2012–2015 Baccalaureate. Mathematics, Lycée National de Mathématiques, Algiers, Algeria. Baccalaureate Mark: 18.31/20 (18-th on the national scale) Other distinctions: Winner of the Regional Mathematics Olympiad.

# Summer and Winter Schools

- Jan. 2022 Winter School e-EGC, Humans in the loop in data mining and learning.
- Oct. 2021 Summer School DSAA, Data-Driven Predictive Maintenance for Industry 4.0.

# Technical Skills

Languages Python, Java, C, C++, C#, Matlab, Bash, Java Script

Frameworks Bootstrap, Microsoft.NET, Spark, Jekyllrb

- Dabatases MySQL, PLSQL, PostgreSQL, HSQL, Oracle, MongoDB
  - ML Sklearn, OpenCV, Keras, TensorFlow, Pytorch, Spacy, NLTK

IDEs VS code, Spider, PyCharm, Netbeans, IntelliJ, Android Studio, Jupyter Lab Others Docker, Git, Github, Postman, Unity3D, Arduino, VHDL

# List of Publications

## National Conferences:

2022 **Youcef Remil**, Anes Bendimerad, Marc Plantevit, Céline Robardet and Mehdi Kaytoue. Découverte de Sous-groupes Interprétables pour le Triage d'incidents. In *Extraction et Gestion des Connaissances, EGC*, vol. 38, pp. 411-418 [Paper, Code, Presentation, Poster]

#### **International Conferences:**

- 2024 Youcef Remil, Anes Bendimerad, Romain Mathonat, Chedy Raissi and Mehdi Kaytoue. DeepLSH: Deep Locality-Sensitive Hash Learning for Fast and Efficient Near-Duplicate Crash Report Detection. In IEEE/ACM International Conference on Software Engineering, ICSE. (Accepted) (Core A\*) [Paper, Code]
- 2023 Anes Bendimerad, **Youcef Remil**, Romain Mathonat and Mehdi Kaytoue. On-premise Infrastructure for AIOps in a Software Editor SME: An Experience Report. In ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering, ESEC/FSE. (Accepted) (Core A\*) [Paper]
- 2021 **Youcef Remil**. How can Subgroup Discovery help AlOps? In *IEEE/ACM International Conference on Automated Software Engineering, ASE*, pp. 1098-1100 (Core A\*) [Paper, Presentation, Poster]
- 2021 Youcef Remil, Anes Bendimerad, Romain Mathonat, Philippe Chaleat and Mehdi Kaytoue. What makes my queries slow? Subgroup Discovery for SQL Workload Analysis. In IEEE/ACM International Conference on Automated Software Engineering, ASE, pp. 642-652 (Core A\*) [Paper, Code, Presentation, Poster]
- 2021 Youcef Remil, Anes Bendimerad, Marc Plantevit, Céline Robardet and Mehdi Kaytoue. Interpretable Summaries of Black Box Incident Triaging with Subgroup Discovery. In IEEE International Conference on Data Science and Advanced Analytics, DSAA, pp. 01-10 (Core A) [Paper, Code, Presentation, Poster]

#### **International Workshops:**

2023 Youcef Remil, Anes Bendimerad, Romain Mathonat, Mathieu Chambard, Marc Plantevit and Mehdi Kaytoue. Subjectively Interesting Subgroups with Hierarchical Targets: Application to Java Memory Analysis. In IEEE International Conference on Data Mining Workshops, ICDMW. (Accepted) [Paper, Code]

### **International Journals:**

2023 **Youcef Remil**, Anes Bendimerad, Romain Mathonat and Mehdi Kaytoue. AlOps solutions for Incident Management: Technical Guidelines and A Comprehensive Literature Review. In *ACM Transactions on Software Engineering and Methodology, TOSEM* (Core A\*). **(To Submit)** 

# Sub-reviewing Activities

- European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECMLPKDD, Core A)
- ACM SIGKDD Conference on Knowledge Discovery and Data Mining (Core A\*)
- The World Wide Web Conference (WWW, Core A\*)
- IEEE International Conference on Data Science and Advanced Analytics (DSAA, Core A)
- Machine Learning and Data Mining for Sports Analytics (MLSA, Workshop ECMLPKDD)

# Contribution to Open Source Projects

- PySubgroup **Subgroup Discovery Package in Python**. **Contribution:** Add new objective interestingness measures such as Support, Mediane, T-Score.
- Mozilla SQL **Parser of SQL queries into JSON format**. Parser **Contribution:** Add a meta-parser for the JSON output to yield a tabular format of SQL queries.

## Internships Supervision

- Avril 2023 4-th year students trinomial, Ecole d'Ingénieurs Informatique EPITA, Lyon, France.
  - Present **Graduation project**: Translating Infologic ERP system from highly specialized and technical French language into multiple diverse languages
    - Conducting an analysis of state-of-the-art multilingual models and machine translation techniques.
    - Fine-tuning the Helsinki model from HuggingFace to ensure accurate translations for English, Spanish, Italian, and Romanian languages.
    - $\circ\;$  Creating a dedicated web portal to facilitate user interaction.
- Jan. 2023 Lynda Belkessa, Ecole Nationale Superieure d'Informatique ESI, Algiers, Algeria.
- June 2023 Graduation project: Time Series Forecasting Methods in the AIOps Domain
  - Exploring time series forecasting and anomaly detection techniques within the domain of AIOps.
  - Implementing the PailoTS framework, which encompasses a wide array of these techniques.
- Sep. 2021 Mathieu Chambard, Ecole Normale Supérieure, Rennes, France.
- Dec. 2021 Internship: Analyzing Java Memory Errors using Subjectively Subgroup Discovery with Hierarchical Target Concepts.
  - Addressing a novel Subjective Subgroup Discovery challenge where the target concept consists of multiple attributes organized in a hierarchy.
  - Conducting experiments with the proposed technique to address many Java memory errors.

# Invited Talks and Seminars

- May 2023 **INRIA Paris**, Representative, Representing the company Infologic in the ANR AT2TA 2023 project (data, challenges, state of the art, etc.).
- Jan. 2023 **EPITA Lyon**, *Journée IA dans l'entreprise*, Speaker, Implementing Preventive Maintenance: From Data Collection to AI in Production.
- July 2022 Université de Lyon 1, Workshop SimpleText@Symposium MADICS, Speaker, Experience on Incident Triage with NLP Techniques in an Industrial Environment [Presentation].
- April 2022 MALOTEC LORIA, Speaker, How Can Subgroup Discovery help AlOps? [Presentation].
- Dec. 2021 LIP6-Sorbonne, Journée Génie Logiciel et Intelligence Artificielle (GLIA), Speaker, Subgroup Discovery for SQL Workload Analysis [Presentation].
- Dec. 2021 KTH Royal Institute of Technology (virtual), Software Technology Research Seminar, Speaker, Subgroup Discovery for SQL Workload Analysis.

## Awards

- Oct. 2021 Winner of Explainable Predictive Maintenance Competition, DSAA Conference 2021.
- Oct. 2018 2-nd Prize Google DevFest Hackathon, Google Developers Group Algiers.
- Nov. 2017 Winner of Arduino Competition, Ecole Nationale Superieure d'Informatique ESI.
- June 2014 Winner of Regional Mathematics Olympiad, Ministry of Education.

# Certification

- March 2023 Advanced SQL Programming, HackerRank.
- March 2023 Intermediate Problem Solving, HackerRank.

- Oct. 2020 Deep Learning Specialization, DeepLearning.Al.
- Oct. 2020 Sequences, Time Series and Prediction, DeepLearning.Al.
- Sep. 2020 Python, HackerRank.
- Sep. 2019 Machine Learning Specialization, Coursera.

## Languages

English Read, Written, Spoken

French Read, Written, Spoken

Arabic Native

## Relevant Educational Projects

March - July 2019
Permutation Flowshop Scheduling Problem.
Technologies: Python, Flask, Plotly
Feb. - July Linux Distribution for Raspberry Pi Embedded System.
Technologies: Linux, LIVE BUILD, Bash
Oct. 2018 - Flow Control Platform based on Qos on a VPN/MPLS Network.
Feb. 2019
Technologies: GNS3, Cisco Packet Tracer
Oct. 2018 - Synchronous Sequence Detector on FPGA Card.
Jan. 2019
Feb. - June Circlab: A Desktop Application for Creating Logic Circuits.
Technologies: C#, .NET

## Miscellaneous Experience

Jan.- Oct. Entrepreneurship Bootcamp, INJAZ EL DJAZAIR (Member of JA Worldwide).
 2018 Experience: Involved in a 9-monthes training course facilitated by renowned consultants in entrepreneurship, business creation, and business models.
 Distinction: Being among the top 10 finalists from all Algerian universities.

July 2018 **Delegate**, International Model United Nations of Middle East and North Africa. **Distinction:** Outstanding Delegate Prize.

## Area of Interests

Motor Sports, Mathematics, Literature and Poems, General Sciences