Master Thesis/Internship positions

Laboratoire d'Informatique Fondamentale et Appliquée
64 avenue Jean Portalis
37200 TOURS
lifat.univ-tours.fr

Équipe Reconnaissance des Formes et Analyse d'Images
www.rfai.li.univ-tours.fr

Machine learning for time series prediction in environmental sciences

Profile: Academic level equivalent to a Master 2 in progress or Engineer in its last year in computer science
Duration: 5 to 6 months, starting from February 2025
Affiliation: Computer Science Lab of the Université de Tours (LIFAT), Pattern Recognition and Image Analysis Group (RFAI)
Salary: according to French legislation (around 620€/month in average); plus indemnity for public transportation, French social help
Supervisor: Nicolas RAGOT, Quentin Besnard (LIFAT-Tours)
Skills:
- a good experience in data analysis and machine learning (in python) is required
- some knowledge and experiences in deep learning and associated tools is required
- some knowledge in time series analysis and forecasting will be highly considered
- curiosity and ability to communicate and share your progress and to make written reports and presentations
- ability to propose solutions
- autonomy and good organization skills

How to candidate:
Send the following documents by e-mail to **nicolas.ragot [at] univ-tours.fr** before 13 of January 2025: a CV, a motivation letter with a short description of projects you worked on and that are related to the topic, your scores including bachelor degree, and references from teachers or people you worked with.

Context:
The JUNON project, driven by the BRGM, is granted from the Centre-Val de Loire region through ARD program (« Ambition Recherche Développement ») which goal is to develop a research & innovation pole around environmental resources (agriculture, forest, waters…). The main goal of JUNON is to elaborate digital services through large scale digital twins in order to improve the monitoring, understanding and prediction of environmental resources evolution and phenomena, for a better management of natural resources. Digital twins will allow to virtually reproduce natural processes and phenomena using combination of AI and environmental tools.
JUNON will focus on the elaboration of digital twins concerning quality and quantity of ground waters, as well as emissions of greenhouse gases and pollutants with health effects, at the scale of geographical area corresponding to the North part of the Centre-Val-de-Loire region.

Nicolas RAGOT
+33 2 47 36 14 31
nicolas.ragot@univ-tours.fr

UNIVERSITÉ DE TOURS
École polytechnique
Département Informatique
64 avenue Jean Portalis
37200 TOURS
polytech.univ-tours.fr

Laboratoire d'Informatique
Fondamentale et Appliquée
64 avenue Jean Portalis
37200 TOURS

lifat.univ-tours.fr

Équipe Reconnaissance des
Formes et Analyse d'Images

www.rfai.li.univ-tours.fr

These digital twins will rely on geological and meteorological knowledge and data (time series), as well as physic-based models.
The project actors are: BRGM, Université d'Orléans, Université de Tours, CNRS, INRAE, and ATOS and ANTEA companies. There are 5 work packages (WP):

1. User's needs and geological knowledge for ground water
2. User's needs and biological/chemical knowledge about pollutants and greenhouse gases
3. Data management and data mining
4. Time series prediction
5. Aggregation and realization of digital twins themselves

The Master Thesis/internship position will be in the WP 4, focused on the prediction of water resources and pollutants in the air. The work will be supervised by the LIFAT - RFAI.

Goals:
The goal of this internship will be to benchmark state of the art time series approaches and to propose new methods adapted to the specificities of the environmental data studied (multivariate time series). The benchmark on water resources relies on complex data with different seasonality and frequencies. Forecasting must be from short term to long term predictions. Regarding air pollutants, the benchmark is still to be elaborated.

Methodology:

1 Data cleaning and preparation will be carefully performed, based on known benchmarks as well as expert knowledge from WP1,2.
2 Classical models from state of the art will be experimented in a rigorous way with a clear report analysis of parameters influence. KAN, PatchMixers, TFT approaches will be particularly studied
3 The generalization of models over multiple series will be investigated as well as the possibility to transfer models.
4 Improvements of the models will be proposed based on either:
- continual learning
- graph representation to model influences between different locations (using GNN)
5 Documentation, reports and cleaning of the code to make it reusable (using Git)

Hosting group:
The RFAI group (Pattern Recognition and Image Analysis) is part of the LIFAT (EA 6300) computer science lab. The group is composed of 3 Professors, 2 HDR (associate professors habilitated), 12 associate or assistant professors, 9 PhDs plus 4 co-supervised PhDs in other universities. The group is working mainly on pattern recognition and machine learning for image/video analysis and temporal data with application domains mainly in health, environment and humanities. The group has access to several computing resources and especially to Leto computer (CPU nodes + 4 gpu nodes with 4 Nvidia Tesla v100 each).

Nicolas RAGOT
+33 2 47 36 14 31
nicolas.ragot@univ-tours.fr

UNIVERSITÉ DE TOURS

École polytechnique
Département Informatique
64 avenue Jean Portalis
37200 TOURS

polytech.univ-tours.fr

Laboratoire d'Informatique
Fondamentale et Appliquée
64 avenue Jean Portalis
37200 TOURS

lifat.univ-tours.fr

Équipe Reconnaissance des
Formes et Analyse d'Images

www.rfai.li.univ-tours.fr

Bibliography:

- G. Zerveas, S. Jayaraman, D. Patel, A. Bhamidipaty, C. Eickhoff. *A Transformer-based Framework for Multivariate Time Series Representation Learning*. arXiv:2010.02803
- Wen, Q., Zhou, T., Zhang, C., Chen, W., Ma, Z., Yan, J., & Sun, L. (2023). *Transformers in Time Series: A Survey*. https://github.com/qingsongedu/time-series-transformers-review
- Casolaro, A., Capone, V., Iannuzzo, G., & Camastra, F. (2023). Deep Learning for Time Series Forecasting: Advances and Open Problems. In *Information (Switzerland)* (Vol. 14, Issue 11). MDPI. https://doi.org/10.3390/info14110598
- Gong, Z., Tang, Y., & Liang, J. (2023). *PatchMixer: A Patch-Mixing Architecture for Long-Term Time Series Forecasting*. http://arxiv.org/abs/2310.00655
- Lim, B., Arik, S. O., Loeff, N., & Pfister, T. (2019). *Temporal Fusion Transformers for Interpretable Multi-horizon Time Series Forecasting*. http://arxiv.org/abs/1912.09363
- Ren, X., Li, X., Ren, K., Song, J., Xu, Z., Deng, K., & Wang, X. (2021). Deep Learning-Based Weather Prediction: A Survey. *Big Data Research*, *23*, 100178. https://doi.org/10.1016/J.BDR.2020.100178
- Tao, H., Hameed, M. M., Marhoon, H. A., Zounemat-Kermani, M., Heddam, S., Sungwon, K., Sulaiman, S. O., Tan, M. L., Sa'adi, Z., Mehr, A. D., Allawi, M. F., Abba, S. I., Zain, J. M., Falah, M. W., Jamei, M., Bokde, N. D., Bayatvarkeshi, M., Al-Mukhtar, M., Bhagat, S. K., … Yaseen, Z. M. (2022). Groundwater level prediction using machine learning models: A comprehensive review. In *Neurocomputing* (Vol. 489, pp. 271–308). Elsevier B.V. https://doi.org/10.1016/j.neucom.2022.03.014
- Uc-Castillo, J. L., Marín-Celestino, A. E., Martínez-Cruz, D. A., Tuxpan-Vargas, J., & Ramos-Leal, J. A. (2023). A systematic review and meta-analysis of groundwater level forecasting with machine learning techniques: Current status and future directions. In *Environmental Modelling and Software* (Vol. 168). Elsevier Ltd. https://doi.org/10.1016/j.envsoft.2023.105788
- Zhang, B., Rong, Y., Yong, R., Qin, D., Li, M., Zou, G., & Pan, J. (2022). Deep learning for air pollutant concentration prediction: A review. *Atmospheric Environment*, *290*, 119347. https://doi.org/10.1016/J.ATMOSENV.2022.119347

Nicolas RAGOT
+33 2 47 36 14 31
nicolas.ragot@univ-tours.fr

UNIVERSITÉ DE TOURS

École polytechnique
Département Informatique
64 avenue Jean Portalis
37200 TOURS

polytech.univ-tours.fr