



PhD-FSTM-2023-061

Faculty of Science, Technology and Medicine

## DISSERTATION

Presented on the 26/06/2023 in Luxembourg

to obtain the degree of

**DOCTEUR DE L'UNIVERSITÉ DU LUXEMBOURG EN  
INFORMATIQUE**

by

Yusuf Arslan

**On the Integration of Interpretable Machine  
Learning Techniques to Machine Learning Pipeline**

### **Dissertation Defense Committee**

Dr. Michail Papadakis, Chair

*Associate Professor, University of Luxembourg, Luxembourg*

Dr. Kevin Allix, Vice-chair

*Assistant Professor, CentraleSupélec, France*

Dr. Jacques Klein, Supervisor

*Full Professor, University of Luxembourg, Luxembourg*

Dr. Christin Seifert, Member

*Full Professor, University of Marburg, Germany*

Dr. Benoît Frénay, Member

*Associate Professor, University of Namur, Belgium*

Dr. Maxime Cordy, Expert

*Research Scientist, University of Luxembourg, Luxembourg*



# Abstract

The long history of machine learning (ML) has led to ML models with super-human results. As a consequence of the superhuman results of ML models, more and more real-world application domains exploit them. However, ML models have disadvantages like opaqueness. In the absence of verification, an opaque black box model is left in the lurch. In addition, when regulations - such as the General Data Protection Regulation in Europe - apply, ML models need to be interpretable. These disadvantages and regulations cause under-use of ML models in highly regulated domains, especially when auditability of processes is demanded, and model understanding is crucial because of high stakes decisions.

Interpretable ML offers a solution to the drawbacks of ML models. It constitutes one of the main instruments that is available to practitioners. If interpretable ML can help practitioners to understand a model better, it could also help algorithms. Indeed, if humans are able to leverage information in explanations of interpretable ML techniques, such information may be methodically exploited in an automated setting. To confirm our intuition, we integrate interpretable ML techniques to the steps of ML pipelines.

This dissertation makes the following contributions to the community:

The first contribution of this thesis is to pinpoint the challenges of interpretable ML in the context of industrial implementation, provide recommendation to overcome these challenges, and thus illuminate the promising research directions as a result of the inspection of an industrial machine learning pipeline.

The second contribution of this thesis is to offer a solution to the issue of ML model performance degradation caused by class imbalance in the resampling step of the model construction stage of the ML pipeline.

The third contribution of this thesis is to highlight the potential of SHAP Explanations to improve model performance, especially for false-positive reduction to decrease the workload of experts while increasing customer trust in industrial settings by inspecting the output of the interpretation step of the post-processing stage of the ML pipeline.

The final contribution of this thesis is to reduce both false-positives and false-negatives by utilizing a second-step classifier based on the SHAP features by using the output of interpretation step of the post-processing stage to obtain a second model for the modeling step of the model construction stage of the ML pipeline.



*The eye of the Sea is one thing and the foam another. Let the foam go, and gaze with the eye of the Sea. Day and night foam-flecks are flung from the sea: oh amazing! You behold the foam but not the Sea. We are like boats dashing together; our eyes are darkened, yet we are in clear water.*

Mevlana Rumi



# Acknowledgements

This work has been supported by the Luxembourg National Research Fund (FNR) under the project ExLiFT (13778825).

This research would not have been possible without the help and support of the many people that I have met throughout my research. I would like to send my deepest gratitude each one of them.

A special note of thanks to all my co-authors who have helped me to develop the results included in this thesis. Besides, the anonymous reviewers of my papers for their insightful comments and suggestions.

I would like to thank BGL BNP Paribas DataLab for granting access to their case studies.









# List of Papers & Services

## Papers included in this dissertation:

- Arslan, Y., Allix, K., Lefebvre, C., Boytsov, A., Bissyandé, T.F. & Klein, J. (2022). Exploiting Prototypical Explanations for Undersampling Imbalanced Datasets. In 21st IEEE International Conference on Machine Learning and Applications (ICMLA), 2022.
- Arslan, Y., Lebichot, B., Allix, K., Veiber, L., Lefebvre, C., Boytsov, A., Goujon, A., Bissyandé, T.F. & Klein, J. (2022). Towards Refined Classifications Driven by SHAP Explanations. In International Cross-Domain Conference for Machine Learning and Knowledge Extraction (CD-MAKE), 2022.
- Arslan, Y., Lebichot, B., Allix, K., Veiber, L., Lefebvre, C., Boytsov, A., Goujon, A., Bissyandé, T.F. & Klein, J. (2022). On the Suitability of SHAP Explanations for Refining Classifications. In Proceedings of the 14th International Conference on Agents and Artificial Intelligence (ICAART), 2022.
- Veiber, L., Allix, K., Arslan, Y., Bissyandé, T. F., & Klein, J. (2020). Challenges Towards Production-Ready Explainable Machine Learning. In 2020 USENIX Conference on Operational Machine Learning (OpML), 2020.

## Papers not included in this dissertation:

- Arslan, Y., Allix, K., Veiber, L., Lothritz, C., Bissyandé, T. F., Klein, J., & Goujon, A. (2021). A comparison of pre-trained language models for multi-class text classification in the financial domain. In Companion Proceedings of the Web Conference, 2021.

## Services:

- 2023 ECML/PKDD Program Committee member
- 2022 ECML/PKDD Program Committee member
- 2021 ECML/PKDD Program Committee member

