The Journal of Open Source Software

pudu: A Python library for agnostic feature selection and explainability of Machine Learning spectroscopic problems.

Enric Grau-Luque[1], **Ignacio Becerril-Romero**[1], **Alejandro Perez-Rodriguez**[1,2], **Maxim Guc**[1], **and Victor Izquierdo-Roca**[1]

1 Catalonia Institute for Energy Research (IREC), Jardins de les Dones de Negre 1, 08930 Sant Adrià de Besòs, Spain **2** Departament d'Enginyeria Electrònica i Biomèdica, IN2UB, Universitat de Barcelona, C/ Martí i Franqués 1, 08028 Barcelona, Spain

DOI: 10.xxxxxx/draft

Software

- Review
- Repository
- Archive

Editor: Open Journals

Reviewers:

- @openjournals

Submitted: 01 January 1970
Published: unpublished

Statement of need

Spectroscopic techniques (e.g. Raman, photoluminescence, reflectance, transmittance, X-ray fluorescence) are an important and widely used resource in different fields of science, such as photovoltaics (Fonoll-Rubio et al., 2022) (Grau-Luque et al., 2021), cancer (Bellisola & Sorio, 2012), superconductors (Fischer et al., 2007), polymers (Easton et al., 2020), corrosion (Haruna et al., 2023), forensics (P. V. Bhatt & Rawtani, 2023), and environmental sciences (Estefany et al., 2023), to name just a few. This is due to the versatile, non-destructive and fast acquisition nature that provides a wide range of material properties, such as composition, morphology, molecular structure, optical and electronic properties. As such, machine learning (ML) has been used to analyze spectral data for several years, elucidating their vast complexity, and uncovering further potential on the information contained within them (Goodacre, 2003) (Luo et al., 2022). Unfortunately, most of these ML analyses lack further interpretation of the derived results due to the complex nature of such algorithms. In this regard, interpreting the results of ML algorithms has become an increasingly important topic, as concerns about the lack of interpretability of these models have grown (Burkart & Huber, 2021). In natural sciences (like materials, physical, chemistry, etc.), as ML becomes more common, this concern has gained especial interest, since results obtained from ML analyses may lack scientific value if they cannot be properly interpreted, which can affect scientific consistency and strongly diminish the significance and confidence in the results, particularly when tackling scientific problems (Roscher et al., 2020).

Even though there are methods and libraries available for explaining different types of algorithms such as SHAP (Lundberg et al., 2017), LIME (Ribeiro et al., 2016), or GradCAM (Selvaraju et al., 2017), they can be difficult to interpret or understand even for data scientists, leading to problems such as miss-interpretation, miss-use and over-trust (Kaur et al., n.d.). Adding this to other human-related issues (Krishnâ1 et al., 2022), researchers with expertise in natural sciences with little or no data science background may face further issues when using such methodologies (Zhong et al., 2022). Furthermore, these types of libraries normally aim for problems composed of image, text, or tabular data, which cannot be associated in a straightforward way with spectroscopic data. On the other hand, time series (TS) data shares similarities with spectroscopy, and while still having specific needs and differences, TS dedicated tools can be a better approach. Unfortunately, despite the existence of several libraries that aim to explain models for TS with the potential to be applied to spectroscopic data, they are mostly designed for a specialized audience, and many are model-specific (Rojat et al., 2021). Furthermore, spectral data typically manifests as an array of peaks that are typically overlapped and can be distinguished by their shape, intensity, and position. Minor shifts in these patterns can indicate significant alterations in the fundamental properties of the subject material. Conversely, pronounced variations in the other case might only indicate negligible differences. Therefore, comprehending such alterations and their implications is paramount. This is still true with ML spectroscopic analysis where the spectral variations are still of primary concern. In this context, a tool with an easy and understandable approach that offers spectroscopy-aimed functionalities that allow to aim for specific patterns, areas, and variations, and that is beginner and non-specialist friendly is of high interest. This can help the different stakeholders to better understand the ML models that they employ and considerably increase the transparency, comprehensibility, and scientific impact of ML results (U. Bhatt et al., 2020) (Belle & Papantonis, 2021).

Overview

pudu is a Python library that quantifies the effect of changes in spectral features over the predictions of ML models and their effect to the target instances. In other words, it perturbates the features in a predictable and deliberate way and evaluates the features based on how the final prediction changes. For this, four main methods are included and defined. **Importance** quantifies the relevance of the features according to the changes in the prediction. Thus, this is measured in probability or target value difference for classification or regression problems, respectively. **Speed** quantifies how fast a prediction changes according to perturbations in the features. For this, the Importance is calculated at different perturbation levels, and a line is fitted to the obtained values and the slope, or the rate of change of Importance, is extracted as the Speed. **Synergy** indicates how features complement each other in terms of prediction change after perturbations. Finally, **Re-activations** account for the number of unit activations in a Convolutional Neural Network (CNN) that after perturbation, the value goes above the original activation criteria. The latter is only applicable for CNNs, but the rest can be applied to any other ML problem, including CNNs. To read in more detail how these techniques work, please refer to the definitions in the documentation.

pudu is versatile as it can analyze classification and regression algorithms for both 1- and 2-dimensional problems, offering plenty of flexibility with parameters, , and the ability to provide localized explanations by selecting specific areas of interest. To illustrate this, Figure 1 shows two analysis instances using the same `importance` method but with different parameters. Additionally, its other functionalities are shown in examples using scikit-learn (Pedregosa et al., 2011), keras (Chollet et al., 2018), and localreg (Marholm, 2022) are found in the documentation, along with XAI methods including LIME and GradCAM.

pudu is built in Python 3 (Van Rossum & Drake, 2009) and uses third-party packages including numpy (Harris et al., 2020), matplotlib (Caswell et al., 2021), and keras. It is available in both PyPI and conda, and comes with complete documentation, including quick start, examples, and contribution guidelines. Source code and documentation are available in https://github.com/pudu-py/pudu.

The Journal of Open Source Software

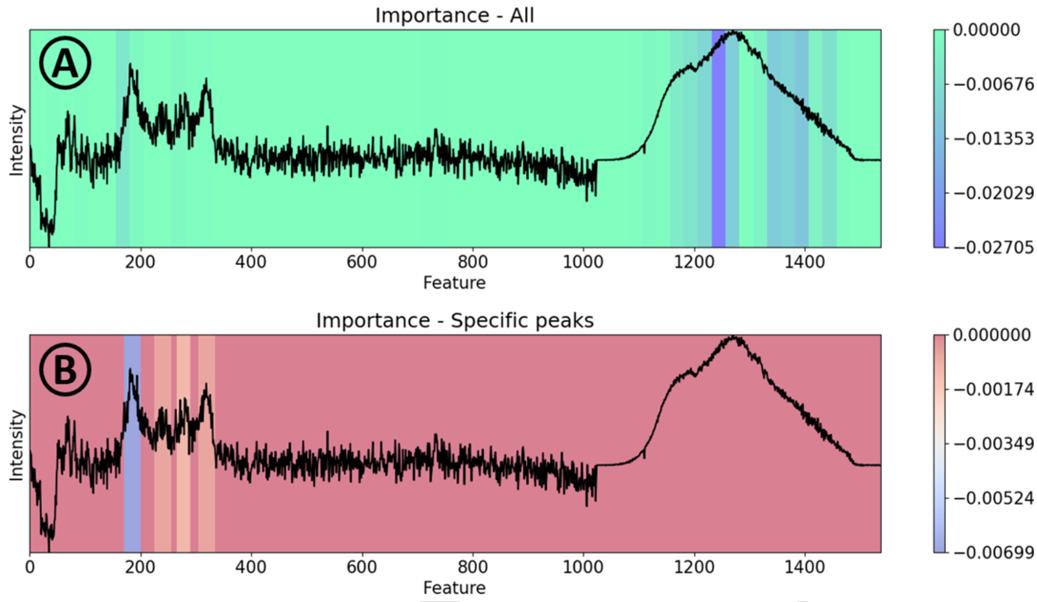

Figure 1: Two ways of using the same method *importance* by A) using a sequential change pattern over all the spectral features and B) selecting peaks of interest. In A), the impact of the peak in the range of 1200-1400 opaques the impact of the rest. In contrast, in B) only the first four main peaks are selected to be analyzed and better visualize their impact in the prediction.

Acknowledgements

Co-funded by the European Union (GA No. 101058459 Platform-ZERO). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Health and Digital Executive Agency (HADEA). Neither the European Union nor the granting authority can be held responsible for them. This project has received funding from the European Union's Horizon 2020 research and innovation programme under Marie Skłodowska-Curie GA No. 801342 (Tecniospring INDUSTRY) and the Government of Catalonia's Agency for Business Competitiveness (ACCIÓ). Authors from IREC belong to the MNT-Solar Consolidated Research Group of the "Generalitat de Catalunya" (ref. 2021 SGR 01286) and are grateful to European Regional Development Funds (ERDF, FEDER Programa Competitivitat de Catalunya 2007–2013).

Authors contribution with CRediT

- Enric Grau-Luque: Conceptualization, Data curation, Software, Writing – original draft
- Ignacio Becerril-Romero: Investigation, Methodology, Writing – review & edition
- Alejandro Perez-Rodriguez: Funding acquisition, Project administration, Resources, Supervision
- Maxim Guc: Formal analysis, Validation, Methodology, Writing – review & edition
- Victor Izquierdo-Roca: Funding acquisition, Project administration, Supervision

References

Belle, V., & Papantonis, I. (2021). Principles and Practice of Explainable Machine Learning. *Frontiers in Big Data*, *4*, 39. https://doi.org/10.3389/FDATA.2021.688969

Bellisola, G., & Sorio, C. (2012). Infrared spectroscopy and microscopy in cancer research and diagnosis. *American Journal of Cancer Research*, *2*(1), 1. /pmc/articles/PMC3236568/

/pmc/articles/PMC3236568/?report=abstract https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3236568/

Bhatt, P. V., & Rawtani, D. (2023). Spectroscopic Analysis Techniques in Forensic Science. *Modern Forensic Tools and Devices: Trends in Criminal Investigation*, 149–197. https://doi.org/10.1002/9781119763406.CH8

Bhatt, U., Xiang, A., Sharma, S., Weller, A., Taly, A., Jia, Y., Ghosh, J., Puri, R., Moura, J. M. F., & Eckersley, P. (2020). *Explainable Machine Learning in Deployment*. https://doi.org/10.1145/3351095.3375624

Burkart, N., & Huber, M. F. (2021). A Survey on the Explainability of Supervised Machine Learning. *Journal of Artificial Intelligence Research*, *70*, 245–317. https://doi.org/10.1613/JAIR.1.12228

Caswell, T. A., Droettboom, M., Lee, A., Andrade, E. S. de, Hunter, J., Hoffmann, T., Firing, E., Klymak, J., Stansby, D., Varoquaux, N., Nielsen, J. H., Root, B., May, R., Elson, P., Seppänen, J. K., Dale, D., Lee, J.-J., McDougall, D., Straw, A., … Ivanov, P. (2021). *matplotlib/matplotlib: REL: v3.4.2*. https://doi.org/10.5281/ZENODO.4743323

Chollet, F., Others, Chollet, F., & Others. (2018). Keras: The Python Deep Learning library. *Astrophysics Source Code Library*, ascl:1806.022. https://ui.adsabs.harvard.edu/abs/2018ascl.soft06022C/abstract

Easton, C. D., Kinnear, C., McArthur, S. L., & Gengenbach, T. R. (2020). Practical guides for x-ray photoelectron spectroscopy: Analysis of polymers. *Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films*, *38*(2). https://doi.org/10.1116/1.5140587/247679

Estefany, C., Sun, Z., Hong, Z., & Du, J. (2023). Raman spectroscopy for profiling physical and chemical properties of atmospheric aerosol particles: A review. *Ecotoxicology and Environmental Safety*, *249*, 114405. https://doi.org/10.1016/J.ECOENV.2022.114405

Fischer, Ø., Kugler, M., Maggio-Aprile, I., Berthod, C., & Renner, C. (2007). Scanning tunneling spectroscopy of high-temperature superconductors. *Reviews of Modern Physics*, *79*(1), 353–419. https://doi.org/10.1103/REVMODPHYS.79.353/FIGURES/62/MEDIUM

Fonoll-Rubio, R., Paetel, S., Grau-Luque, E., Becerril-Romero, I., Mayer, R., Pérez-Rodríguez, A., Guc, M., & Izquierdo-Roca, V. (2022). Insights into the Effects of RbF-Post-Deposition Treatments on the Absorber Surface of High Efficiency Cu(In,Ga)Se2 Solar Cells and Development of Analytical and Machine Learning Process Monitoring Methodologies Based on Combinatorial Analysis. *Advanced Energy Materials*, 2103163. https://doi.org/10.1002/AENM.202103163

Goodacre, R. (2003). Explanatory analysis of spectroscopic data using machine learning of simple, interpretable rules. *Vibrational Spectroscopy*, *32*(1), 33–45. https://doi.org/10.1016/S0924-2031(03)00045-6

Grau-Luque, E., Anefnaf, I., Benhaddou, N., Fonoll-Rubio, R., Becerril-Romero, I., Aazou, S., Saucedo, E., Sekkat, Z., Perez-Rodriguez, A., Izquierdo-Roca, V., & Guc, M. (2021). Combinatorial and machine learning approaches for the analysis of Cu2ZnGeSe4: influence of the off-stoichiometry on defect formation and solar cell performance. *Journal of Materials Chemistry A*, *9*(16), 10466–10476. https://doi.org/10.1039/d1ta01299a

Harris, C. R., Millman, K. J., Walt, S. J. van der, Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., Kerkwijk, M. H. van, Brett, M., Haldane, A., Río, J. F. del, Wiebe, M., Peterson, P., … Oliphant, T. E. (2020). *Array programming with NumPy* (No. 7825; Vol. 585, pp. 357–362). Nature Research. https://doi.org/10.1038/s41586-020-2649-2

Haruna, K., Obot, I. B., & Saleh, T. A. (2023). Infrared Spectroscopy in Corrosion Research. *Corrosion Science*, 261–289. https://doi.org/10.1201/9781003328513-9

Kaur, H., Nori, H., Jenkins, S., Caruana, R., Wallach, H., & Wortman Vaughan, J. (n.d.). *Interpreting Interpretability: Understanding Data Scientists' Use of Interpretability Tools for Machine Learning*. https://doi.org/10.1145/3313831.3376219

Krishnå1, S., Han˚1, T. H., Gu, A., Pombra, J., Jabbari, S., Wu, Z. S., & Lakkaraju, H. (2022). *The Disagreement Problem in Explainable Machine Learning: A Practitioner's Perspective*. https://arxiv.org/abs/2202.01602v3

Lundberg, S. M., Allen, P. G., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. *Advances in Neural Information Processing Systems*, *30*. https://github.com/slundberg/shap

Luo, R., Popp, J., & Bocklitz, T. (2022). Deep Learning for Raman Spectroscopy: A Review. *Analytica*, *3*(3), 287–301. https://doi.org/10.3390/analytica3030020

Marholm, S. (2022). *sigvaldm/localreg: Multivariate RBF output*. https://doi.org/10.5281/ZENODO.6344451

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). *Scikit-learn: Machine Learning in Python* (Vol. 12, pp. 2825–2830). http://scikit-learn.sourceforge.net.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why Should I Trust You?": Explaining the Predictions of Any Classifier. *NAACL-HLT 2016 - 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Proceedings of the Demonstrations Session*, 97–101. https://doi.org/10.48550/arxiv.1602.04938

Rojat, T., Puget, R., Filliat, D., Del Ser, J., Gelin, R., & Díaz-Rodríguez, N. (2021). *Explainable Artificial Intelligence (XAI) on TimeSeries Data: A Survey*. https://arxiv.org/abs/2104.00950v1

Roscher, R., Bohn, B., Duarte, M. F., & Garcke, J. (2020). Explainable Machine Learning for Scientific Insights and Discoveries. *IEEE Access*, *8*, 42200–42216. https://doi.org/10.1109/ACCESS.2020.2976199

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual Explanations From Deep Networks via Gradient-Based Localization. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 618–626). https://doi.org/10.1109/iccv.2017.74

Van Rossum, G., & Drake, F. L. (2009). Python 3 Reference Manual; CreateSpace. *Scotts Valley, CA*, 242. ISBN: 978-1-4414-1269-0

Zhong, X., Gallagher, B., Liu, S., Kailkhura, B., Hiszpanski, A., & Han, T. Y. J. (2022). Explainable machine learning in materials science. *Npj Computational Materials 2022 8:1*, *8*(1), 1–19. https://doi.org/10.1038/s41524-022-00884-7