FAIRNESS ASSESSMENT AND BIAS MITIGATION IN MACHINE LEARNING

Dr. Eufrasio de Andrade Lima Neto

ABOUT THE SPEAKER

Eufrásio de A. Lima Neto is an Associate Professor in the Department of Statistics and a Faculty Member of the Graduate Program in Decision Models and Health at the Federal University of Paraíba (Brazil). He was a Research Associate in Data Science in the Department of Computer Science at Loughborough University (UK, 2022) and a Senior Lecturer in Data Analytics at De Montfort University (UK, 2023–2024). His main research interests include statistical modeling, generalized linear models, robust regression, clusterwise regression, symbolic data analysis, interval-valued data, kernel methods, machine learning, and fairness assessment in ML algorithms. He is the author of over 50 technical papers published in international journals and conference proceedings. He is an ISI Elected Member and has held various leadership roles, including serving on the Board of Directors of LARS-IASC (2025–2028, 2017–2020), as Executive Secretary of the Brazilian Association of Statistics (ABE, 2018–2020), and as a Member of the Board of Directors of ABE (2020–2024).

Course: Fairness Assessment and Bias Mitigation in Machine Learning (4 hours)
Instructor: Dr. Eufrasio de Andrade Lima Neto

Objective and Program

This course introduces key concepts in ethics, fairness assessment, and bias mitigation for machine learning algorithms applied to classification problems. Participants will explore metrics for detecting algorithmic bias and mitigation strategies, all grounded in real-world datasets and supported by practical implementation in R and/or Python. The session is designed to ensure a clear and accessible understanding of the subject.

Day 1 (2 hours)
- Introduction to ethics and fairness assessment
- Identifying fairness-related harms
- Measuring and detecting bias (metrics)
- Bias mitigation techniques
- Bias mitigation algorithms
- Relevant libraries and packages

Day 2 (2 hours)
- The fairmodels library in R
- Hands-on: German credit data example
- The fairlearn package in Python (Google Collab)
- Hands-on: Healthcare example

Materials and Readings
- Course's Slides
- R script
- Python Notebook
- https://fairlearn.org/
- Wiśniewski J. and Biecek, P. fairmodels: A Flexible Tool For Bias Detection, Visualization, And Mitigation
(available on arxiv: https://arxiv.org/abs/2104.00507)
- Ntoutsi E, Fafalios P, Gadiraju U, et al. Bias in data-driven artificial intelligence systems—An introductory survey. WIREs Data Mining Knowl Discov. 2020; 10:e1356. https://doi.org/10.1002/widm.1356
- Neto, E.d.A.L., Bailiss, J., Finke, A. et al. Identifying early help referrals for local authorities with machine learning and bias analysis. J Comput Soc Sc 7, 385–403 (2024). https://doi.org/10.1007/s42001-023-00242-7

ShortBio

Eufrásio de A. Lima Neto is an Associate Professor in the Department of Statistics and a Faculty Member of the Graduate Program in Decision Models and Health at the Federal University of Paraíba (Brazil). He was a Research Associate in Data Science in the Department of Computer Science at Loughborough University (UK, 2022) and a Senior Lecturer in Data Analytics at De Montfort University (UK, 2023–2024). His main research interests include statistical modeling, generalized linear models, robust regression, clusterwise regression, symbolic data analysis, interval-valued data, kernel methods, machine learning, and fairness assessment in ML algorithms. He is the author of over 50 technical papers published in international journals and conference proceedings. He is an ISI Elected Member and has held various leadership roles, including serving on the Board of Directors of LARS-IASC (2025–2028, 2017–2020), as Executive Secretary of the Brazilian Association of Statistics (ABE, 2018–2020), and as a Member of the Board of Directors of ABE (2020–2024).