S.A.F.E. metrics for high quality artificial Intelligence in finance

Paolo Giudici*, Golnoosh Babaei, Emanuela Raffinetti

University of Pavia

Extended Abstract

The growth of Artificial Intelligence (AI) applications requires to develop quality management models. In this paper, we contribute to the debate on quality standards for artificial intelligence applications proposing a set of integrated SAFE - AI statistical metrics. Our proposed metrics are consistent with each other, as they are all derived from a common underlying statistical methodology: the Lorenz curve and the related Gini index. They are easy to interpret, as are all expressed in percentages of an ideal situation of full quality. They are agnostic, as they can be applied to any machine learning method. They are fully reproducible, using the proposed Python code. Our SAFE-AI methodological framework allows to derive any necessary quality metric for AI using a pairwise comparison of Lorenz curves, based on different rank graduations. We specifically consider metrics for the assessment of Sustainability, Accuracy, Fairness, and Explainability. Our proposal allows the application of all the proposed metrics for the assessment of the quality of any AI applications, in finance, but also in other sectors such as health care and manufacturing.

In the talk we have reviewed some recent papers of ours that have built the SAFE-AI framework, as follows.

The paper Giudici and Kolesnikov (2026) proposes how to integrate the measures in a single measurement; the paper Babaei et al. (2025) specifies the metrics; the paper Giudici and Raffinetti (2025) constructs the base ingredient, the Rank Graduation Accuracy (RGA) measure; the papers Auricchio et al. (2025) and Giudici et al. (2025) describe how to extend the metrics to the multidimensional case; the paper by Calzarossa et al. (2025) focuses on the explainability metric, whereas Chen et al. (2024) on the fairness metric. On the application side, Giudici et al. (2024) consider financial time series; and Babaei et al. (2024) generative AI; Giudici et al. (2024) sets the stage for an AI risk management framework.

*= Presenter

References

Giudici, P., Kolesnikov, V. (2026) . SAFE AI metrics: an integrated approach. Machine Learning with Applications, 23, 100821

Babaei, G., Giudici, P., Raffinetti, E. (2025). A Rank graduation box for SAFE AI. Expert systems with applications, 59, 125239.

Giudici, P., Raffinetti, E. (2025). RGA: a unified measure of predictive accuracy. Advances in Data Analysis and Classification, 19(1), 67-93,114104.

G. Auricchio, P. Giudici, G. Toscani, (2025) Extending the Gini index to higher dimensions using whitening processes. Rend. Linc. Mat. Appl. 35 (3), pp. 511–528

Giudici, P., Raffinetti, E., Toscani, G. (2025) Measuring multidimensional inequality: a new proposal based on the Fourier transform. Statistics, 2025, 59(2), 330-353.

Calzarossa, Giudici, Zieni (2025). An assessment framework for explainable AI, with applications to cybersecurity. Artificial Intelligence review, 2025.

Chen, Y., Giudici, P., Kailang, L., Raffinetti, E. (2024) Measuring fairness in credit ratings. Expert systems with applications, 258, 125184.

Giudici, Piergallini, Raffinetti, Recchioni. Explainable AI methods for financial time series. Physica A: statistical mechanics with applications, 2024, 655, 130176.

Babaei, G., Giudici, P. (2024). GPT classification, with application to lending. Machine learning with applications, 16, 100534.

Giudici, P., Centurelli, M., Turchetti, S. (2024) Artificial Intelligence risk measurement. Expert systems with applications, 235, 121220.