

Hi Michael,

Imagine someone with a high income but a limited credit history. SHAP might reveal that the model is giving too much weight to credit history, which could lead to an unfair rejection. That kind of breakdown is helpful for improving model fairness.

I also like your point on feature importance. Can SHAP flag when a model over-relies on one factor, like credit history, while downplaying others, like job stability?

You raised an important concern about multicollinearity. It made me wonder: What if we combine related features into a single composite score, such as bundling income, payment history, and credit usage, into one financial behavior score? Would SHAP still offer useful insights, or might we lose too much detail (Hjelkrem & de Lange, 2023)? Have you experimented with techniques like PCA or feature grouping for this (Vieira, 2021)? Great work presenting both strengths and limitations of SHAP so clearly.

All The Best,

Avinash Bunga

References

Hjelkrem, L. O., & de Lange, P. E. (2023). Explaining deep learning models for credit scoring with SHAP: A case study using open banking data. *Journal of Risk and Financial Management*, 16(4), 221. <https://doi.org/10.3390/jrfm16040221>

Vieira, E. U. P. (2021, April 2). *You are underutilizing SHAP values — Feature groups and correlations*. Medium. Retrieved April 12, 2025, from <https://medium.com/data-science/you-are-underutilizing-shap-values-understanding-populations-and-events-7f4a45202d5>