Report on ”Beyond Generic Phishing Detection: Explainable AI for Finance-Adapted Models in Banking and Fintech”

The manuscript addresses an interesting and practically relevant question about finance-adapted phishing detection across email and webpage settings. The multi-modal framing, model comparison, and use of explainability tools make the paper potentially useful. I am generally supportive of publication, but a few focused clarifications would help keep the final claims aligned with the evidence.

1. The finance-theme labelling strategy is central to the contribution and should be more transparent. Because the FT/NFT label determines the finance-adapted training set, subgroup evaluation, and interpretation of domain-specific gains, readers need enough detail to assess this construction. The authors should provide the full dictionary or appendix-level details, examples of FT and NFT records, and a small manual validation exercise or sensitivity check around the threshold and weights. It would also be useful to clarify whether the finance terms used for labelling overlap with the TF-IDF features used by the email classifiers.

2. Several sample-count and subgroup definitions need to be reconciled. The data section describes the email dataset as 82,486 observations, while the abstract and Table 21 refer to 164,972 email instances. If this reflects pooling, duplication, augmentation, or preprocessing, the explanation should be explicit. The webpage analysis also needs clarification: Table 21 reports zero finance-themed webpage observations, yet Tables 18 and 19 report FT webpage results with N = 321. The bootstrap gains in Figure 5 also appear larger than the corresponding F1 differences in Tables 16 and 17. A compact audit table with sample counts and gain calculations by modality would help.

3. The interpretation should be framed as promising but conditional. The evidence is strongest for email models, while the webpage results appear to rely mainly on broad structural and reputation features. The XAI and practical discussion should therefore avoid implying strong finance-specific webpage learning unless the FT webpage subset is clearly established. The discussion may also benefit from briefly situating the fintech platform context. For example, Bao and Huang (2021) may be useful for positioning fintech platforms as digital financial-intermediation settings where risk management and model deployment choices have practical significance.