Peer Review

Review of: "An Empirical Analysis of Loan Repayment Behavior and Default Rates on Digital Lending Platforms: Evidence from an Emerging Market"

Rakshith Bhandary[1]

1. Manipal University, India

This paper examines the input variables of loan repayment behavior on digital lending platforms. The idea is interesting, and the method employed is appropriate for the large corpus of literature in the field. Nevertheless, the paper could be improved in the following ways, as per the suggestions given below.

1. The title does not look appropriate for the work done in this study. "*An Empirical Analysis of Loan Repayment Behavior and Default Rates on Digital Lending Platforms: Evidence from an Emerging Market*" is the mentioned title. Since the study does not calculate the default rates in any manner, I suggest removing it. Since the variables that affect default behaviour are tested, it is advised to rework and rename the title. I suggest "Predictors of loan default behaviour on digital lending platforms: empirical evidence from Kenya"
2. In the Introduction section, please authenticate the statements made with proper citation. For example: *"With the advent of mobile-based digital lending platforms, individuals previously excluded from formal financial systems now have unprecedented access to credit"* – even though the statement is generally accepted, citing the source to substantiate it will improve credibility. You can cite the economic survey, articles, or central bank data or prominent research organizations' data. Please validate the statements throughout the introduction using the source citation method. The introduction can be improved by highlighting the importance of loan default prediction according to the global convention held by the BIS on capital adequacy ratios, risk weightage, Basel norms, credit default risk, expected loss, probability of default as per the IFRS 9 standards,

etc. Highlight the need for this study and end with research questions (either at the end of the introduction or in the literature review)

3. The literature can highlight the input variables used in loan default prediction. Income and loan amount must be written here as they are the significant predictors of default as per the majority of the studies worldwide (I don't want to give references as it may cause bias, but please check online for sources that show the significant relationship of income and loan amount with default). Also, these variables (income and loan amount) are practically used in industrial applications (even though this study mentions "*monthly income, loan amount, and awareness of credit scores, were initially considered but excluded from the final model due to multicollinearity or statistical insignificance*"- strong justification in the discussion section for excluding these variables in this study is required). The hypotheses formulation must be strengthened (it is better to support it with theory wherever possible). The first hypothesis, for example, the relationship between age and default, must be referenced with relevant literature. H1 is the alternative hypothesis, and it is not written properly (it should have been written as *The age of a borrower on digital platforms significantly affects the likelihood of loan default*); also, rewrite the note on hypotheses 4.4.5. The same applies to H2 and H4. Hypotheses formulation is weak without sufficient literature.
4. The study mentions that the data was collected through a structured questionnaire (please cite the source from which the questionnaire was adopted for each variable). Also, please mention how the qualitative information was captured (were open-ended questions asked? or were interviews conducted? How was it analysed?). Was the content validated by experts in the digital lending domain before pre-testing with 10 respondents? Pilot testing with 30 samples could have been done. The sample size was just 161 (usually, 400 is desired for an infinite population); please justify based on the population (use G-power software or any other, or based on the thumb rule for the structured scale items in the questionnaire). Please mention how you tackled common method bias in social science research (pre-measures and post-measures like the Harmon single factor or the correlation marker approach). Please justify the use of the probit model for analysis (why not the logit or any other machine learning methods like SVM, logistic regression, tree-based algorithm, or ANN, etc.). Even though a robustness check and multicollinearity were performed, additional validity measures are desirable (not mandatory but desirable).
5. The Discussion section should compare the results and findings of this study with previous studies as per the literature. Similarities in findings and/or contradictions should be defended,

doi.org/10.32388/BPHO04

justified, and substantiated here in this section.

6. The Conclusion should include key takeaways and the gist of the study, starting with the main objective, findings, implications, limitations, and the scope for future studies must be written here. The conclusion section can be written in two sections with key takeaways from this study, limitations, and directions for future research. A brief point on the practical implication of this research can also be written (implications for managers and practitioners).
7. The English language looks fine. Additional references must be added to strengthen the paper. References quoted should mean the same as written in the manuscript. Kindly cross-check the entire reference list for accuracy and kindly use source citation in each section (Cite the source).

Declarations

Potential competing interests: No potential competing interests to declare.