

LightGBM-Based Prediction of Marital Satisfaction Using Emotional, Sexual, and Communication Indicators

Ingrid. Løvik¹, Thomas. Campbell², Elin. Sjöberg³

¹ Department of Social Psychology, University of Oslo, Oslo, Norway

² Department of Experimental Psychology, University of New Brunswick, Fredericton, Canada

³ Department of Developmental Psychology, Lund University, Lund, Sweden

* Corresponding author email address: tcampbel@unb.ca

E d i t o r	R e v i e w e r s
Monika Szczygieł  Department of Psychology, Jagiellonian University, Krakow, Poland monika.szczygieł@uj.edu.pl	Reviewer 1: Mohammadreza Zarbakhsh Bahi  Associate Professor (Department of Psychology, Tonekabon Branch, Islamic Azad University, Tonekabon, Iran. Email: M.Zarbakhsh@Toniau.ac.ir Reviewer 2: Fahime Bahonar  Department of counseling, University of Isfahan, Isfahan, Iran. Email: Fahime.bahonar@edu.ui.ac.ir

1. Round 1

1.1. Reviewer 1

Reviewer:

The Study Design and Participants section reports that participants were required to be “currently legally married or living in a marital relationship recognized as equivalent for the purposes of the study,” yet the title, terminology, and interpretation throughout the paper consistently refer to “marital satisfaction.” This operational definition requires clarification. Legally married participants and individuals in marriage-equivalent cohabiting unions may differ in relationship duration, legal commitments, family structure, and sociocultural expectations. The authors should either restrict the analytic population to legally married adults, revise terminology to “relationship satisfaction,” or report the proportions in each partnership category and demonstrate that the results are robust after adjusting for legal marital status. The current wording introduces a conceptual mismatch between the target construct, study population, and claimed scope of inference.

In the Data Collection Tools section, the manuscript states that marital satisfaction was assessed using the “10-item Couples Satisfaction Index (CSI-10).” The authors should verify the instrument nomenclature and provide a precise scoring description. The widely cited Couples Satisfaction Index has established versions and specific item response formats, and therefore the exact version used, possible total-score range, scoring procedure, treatment of reverse-coded items, and evidence supporting

its use in Canadian populations should be clearly documented. If the manuscript uses a modified 10-item form, the authors must explain how that form was derived and provide psychometric justification. Because the dependent variable determines all subsequent performance metrics, uncertainty regarding the outcome's construction directly affects interpretation of R^2 , RMSE, MAE, and the clinical meaning of prediction errors.

Table 2 reports “Cross-Validated R^2 , Mean $\pm$ SD” alongside independent-test R^2 , RMSE, and MAE, but statistical uncertainty around the comparative metrics is absent. The authors should estimate 95% confidence intervals for test-set R^2 , RMSE, and MAE, preferably using appropriate bootstrap procedures applied to the held-out predictions. In addition, pairwise comparison of prediction errors should be considered to determine whether LightGBM meaningfully outperformed XGBoost and random forest rather than merely producing numerically better point estimates. The present wording risks overstating algorithmic superiority from relatively small differences. The results should distinguish “highest observed performance” from evidence of statistically or practically meaningful improvement.

Table 3 presents highly specific optimized values such as “436 boosting iterations, learning rate = 0.031, number of leaves = 27, L1 regularization = 0.18, and L2 regularization = 0.74.” These values imply a formal hyperparameter-search procedure, but the manuscript does not state whether grid search, random search, Bayesian optimization, Optuna, or another method was used. The authors should report the complete search ranges or prior distributions for every tuned parameter, the objective function used to choose the optimum, the number of evaluated configurations, and whether early stopping was applied inside each cross-validation fold. Otherwise, these precise parameter values are not reproducible and could appear retrospectively selected to match the reported performance.

Response: Revised and uploaded the manuscript.

1.2. Reviewer 2

Reviewer:

A major measurement concern appears in the paragraph describing “perceived emotional support from the spouse,” which states that this construct was assessed “through items assessing the extent to which participants experienced understanding, emotional availability, validation, and responsiveness.” Unlike the named standardized instruments, no scale name, item count source, validation reference, response format, or development procedure is provided. A similar issue occurs for “sexual communication,” “sexual-need compatibility,” and some communication variables. If these are researcher-developed measures, the authors should explicitly identify them as such and report item generation, expert review, content-validity procedures, dimensionality assessment, reliability, and preferably construct-validity evidence. Predictive models can exploit measurement artifacts, so inadequately validated predictors may yield strong feature importance without necessarily representing stable psychological constructs.

In the Data Collection Tools section, the manuscript simultaneously proposes use of the “New Sexual Satisfaction Scale–Short Form (NSSS-S),” additional sexual communication indicators, sexual-need compatibility, and sexual intimacy frequency. The conceptual overlap among these variables needs to be evaluated explicitly. Although multicollinearity does not destabilize tree models in the same manner as ordinary least-squares regression, correlated predictors can substantially distort feature-importance rankings because importance may be divided among substitutable variables or attributed preferentially to one correlated feature. The authors should report intercorrelations among predictors, identify highly redundant features, and conduct a stability analysis of SHAP importance rankings under removal of correlated variables. Without this assessment, the conclusion that sexual satisfaction was the “most influential” predictor may partly reflect how the feature set was operationalized.

The manuscript states in the Data Analysis section that missing values were handled using “appropriate imputation procedures based exclusively on the training data to prevent information leakage.” This is methodologically appropriate in principle but too vague for reproducibility. The authors should report the percentage of missingness for every analyzed variable, define the threshold used to exclude cases for “extensive missing data,” identify the exact imputation algorithm applied to

continuous and categorical features, and specify whether imputation occurred separately within each cross-validation fold rather than once on the entire training set. The latter distinction is critical: fitting an imputer to the full training sample before cross-validation can leak information from validation folds into model development and yield optimistic validation estimates.

The Data Analysis paragraph reports an “80:20 allocation,” yielding 369 training and 93 test participants. Given the modest total sample size and the relatively large set of candidate predictors, a single random holdout split may produce an unstable estimate of generalization performance. The manuscript should provide the random seed and explain whether the train-test split was stratified according to the distribution of marital satisfaction or other relevant characteristics. More importantly, the authors should supplement the single holdout analysis with repeated nested cross-validation or repeated train-test splitting to quantify uncertainty in R^2 , RMSE, and MAE. Confidence intervals around test performance would be especially valuable because differences such as $R^2 = .772$ for LightGBM versus $.739$ for XGBoost may not represent a reliable performance advantage in only 93 test observations.

The manuscript claims that “LightGBM demonstrated the strongest overall out-of-sample performance” based on comparisons with multiple linear regression, Elastic Net, random forest, and XGBoost. However, the fairness of this comparison cannot be evaluated from the information provided. The authors should specify whether each comparator received equivalent hyperparameter tuning, identical cross-validation folds, identical preprocessing, and exactly the same predictor set. Random forest and XGBoost, in particular, require meaningful tuning to provide a credible benchmark. A heavily optimized LightGBM model compared with default or minimally tuned alternatives would bias the comparison. The manuscript should therefore provide the tuning spaces, optimization strategy, selection criterion, and computational budget for every algorithm.

Response: Revised and uploaded the manuscript.

2. Revised

Editor’s decision after revisions: Accepted.

Editor in Chief’s decision: Accepted.