Personalized Recommendation and Consumer Conversion: Evidence from Commercial Advertising Data

Tao Luo[1], Ruyi Yin[2], Yang Guo[3]*

[1]College of Earth and Environmental Sciences, Lanzhou University, Lanzhou, 730100, China
[2]Chongqing Engineering Vocational College, Chongqing, 402260, China
[3]College of Economics and Management, Harbin Engineering University, Harbin, 150001, China

Keywords:

Personalization;
Recommendation systems;
Advertising;
Conversion prediction;
DeepFM;
Explainable AIAIGC;
Traditional culture

ABSTRACT

This study examines how AI-driven personalized recommendation relates to consumer conversion in digital advertising, using the Criteo commercial advertising dataset. Descriptive statistics, Pearson correlation, logistic regression, and DeepFM are used to analyze conversion behavior and compare predictive performance. The study reports an 11.6% conversion rate and finds that a one-unit increase in AI recommendation Match is associated with approximately 4.4% higher conversion odds ($\beta = 0.0431$, $p < .01$). DeepFM achieves an AUC of 0.8471 and a LogLoss of 0.3102, outperforming logistic regression (AUC = 0.6269) and random forest (AUC = 0.7249) on the principal discrimination and loss measures; its accuracy is 0.8842. Robustness analyses include an alternative CTR measure, mobile-user subsample, Probit estimation, and propensity-score matching. The study reports a positive PSM average treatment effect of 0.0427 ($p < .01$) and concludes that personalized matching is an important factor associated with conversion, while user activity and advertising context also contribute to conversion outcomes. The findings support the use of DeepFM for conversion prediction and motivate dynamic recommendation weighting and coordinated marketing optimization.

CONTACT:

[1]Author: luot2024@lzu.edu.cn

*Corresponding Author: 13861396887@hrbeu.edu.cn

Note: Authors designated with " † " are recognised as co-first authors.

DOI.10.64549/jaai-ii.v1i2.122

This work is licensed under the CC BY 4.0 HTTPS://CREATIVECOMMONS.ORG/LICENSES/BY/4.0/

1. Introduction

Digital advertising systems use behavioral and contextual information to select content and estimate the probability of a click or purchase. Recommendation models represent interactions among users, products, devices, and contexts, supporting personalized ad delivery and conversion prediction. This study examines the relationship between personalized matching, user activity, and consumer conversion through logistic regression and DeepFM, and evaluates how these approaches inform conversion-oriented marketing.

The baseline results show that a one-unit increase in the Match score is associated with approximately 4.4% higher conversion odds, and DeepFM achieves an AUC of 0.8471. The study further examines alternative feature measurement, model specification, and mobile-user scenarios. Across these analyses, the authors conclude that personalized matching is positively related to conversion and that DeepFM provides strong predictive performance.

2. Literature and hypotheses

Recommender-system research has progressed from collaborative filtering to latent-factor and feature-interaction models, with DeepFM and related architectures designed to learn nonlinear relationships in sparse behavioral data (Resnick & Varian, 1997; Adomavicius & Tuzhilin, 2005; Sarwar et al., 2001; Rendle, 2010; Guo, 2017; Cheng, 2016; Covington et al., 2016; He & Chua, 2017; McMahan, 2013; Ricci et al., 2011). Large-scale recommendation and advertising studies describe applications of these models to user response prediction (Chapelle et al., 2014; Criteo AI Lab, 2014). Building on this work, the present study compares an interpretable logistic baseline with a nonlinear predictor to examine personalized matching and consumer conversion.

Prior empirical research applies online experimentation, propensity-score methods, and alternative statistical specifications to evaluate recommendation and advertising outcomes (Deng, 2013; Rosenbaum & Rubin, 1983). Predictive alternatives such as boosting and random forests provide complementary benchmarks, while ridge, Lasso, and quantile regression offer additional approaches to modeling (Breiman, 2001; Chen & Guestrin, 2016; Friedman, 2001; Hoerl & Kennard, 1970; Koenker & Bassett, 1978; Tibshirani, 1996). Related work on mediation and panel estimation provides further methodological tools for empirical analysis (Baron & Kenny, 1986; Arellano & Bond, 1991). These approaches inform the model comparison and robustness analyses in this study.

Interpretability and responsible deployment form a third strand. SHAP-style attribution and local explanation methods describe how fitted models use their inputs (Lundberg & Lee, 2017; Ribeiro et al., 2016). Privacy law, anonymization research, and cooperative-game theory inform the treatment of behavioral data and the interpretation of feature contributions (European Union, 2016; Sweeney, 2002; Shapley, 1953). Work on concept drift and technical debt highlights the value of reproducible preprocessing and ongoing model monitoring (Gama, 2014; Sculley, 2015; Cao, 2017). These perspectives complement the study's analysis of personalized matching, user activity, and advertising context using the Criteo commercial advertising dataset.

The study proposes that recommendation relevance is positively associated with conversion (H1), and that models capturing nonlinear feature interactions improve predictive performance (H2). Table 2 summarizes the model evaluation design. User activity and ad position are also examined as factors associated with conversion. These hypotheses guide the empirical analysis of personalized matching and conversion prediction using logistic regression and DeepFM.

3. Data and empirical strategy

The study uses a Criteo commercial advertising sample described as containing 100,000 observations, 13 continuous variables, and 26 categorical variables. The source reports an 11.6% conversion rate and identifies 62,234 training observations and 15,559 test observations for model estimation and evaluation. The feature set covers user behavior, product information, advertising context, device type, and visit time.

The baseline logistic model is

$$Pr(Convert_i = 1) = \Lambda(\alpha + \beta Match_i + \theta Activity_i + \delta' Z_i) \quad (1)$$

where σ is the logistic link and Z includes the advertising and device controls. The source interprets exp(β) as the odds ratio. DeepFM and conventional models are compared using AUC, accuracy, and log loss. The analysis also uses model attribution methods to summarize the predictive contribution of the input features (Lundberg & Lee, 2017; Ribeiro et al., 2016).

The empirical strategy combines an alternative measure of the core predictor, a mobile-user subsample, Probit estimation, and propensity-score matching. These analyses examine the reported associations across alternative measures, estimation approaches, and user scenarios. Figure 1 summarizes the research workflow.

The empirical sequence from source review through model estimation and validation is summarized in Figure 1.

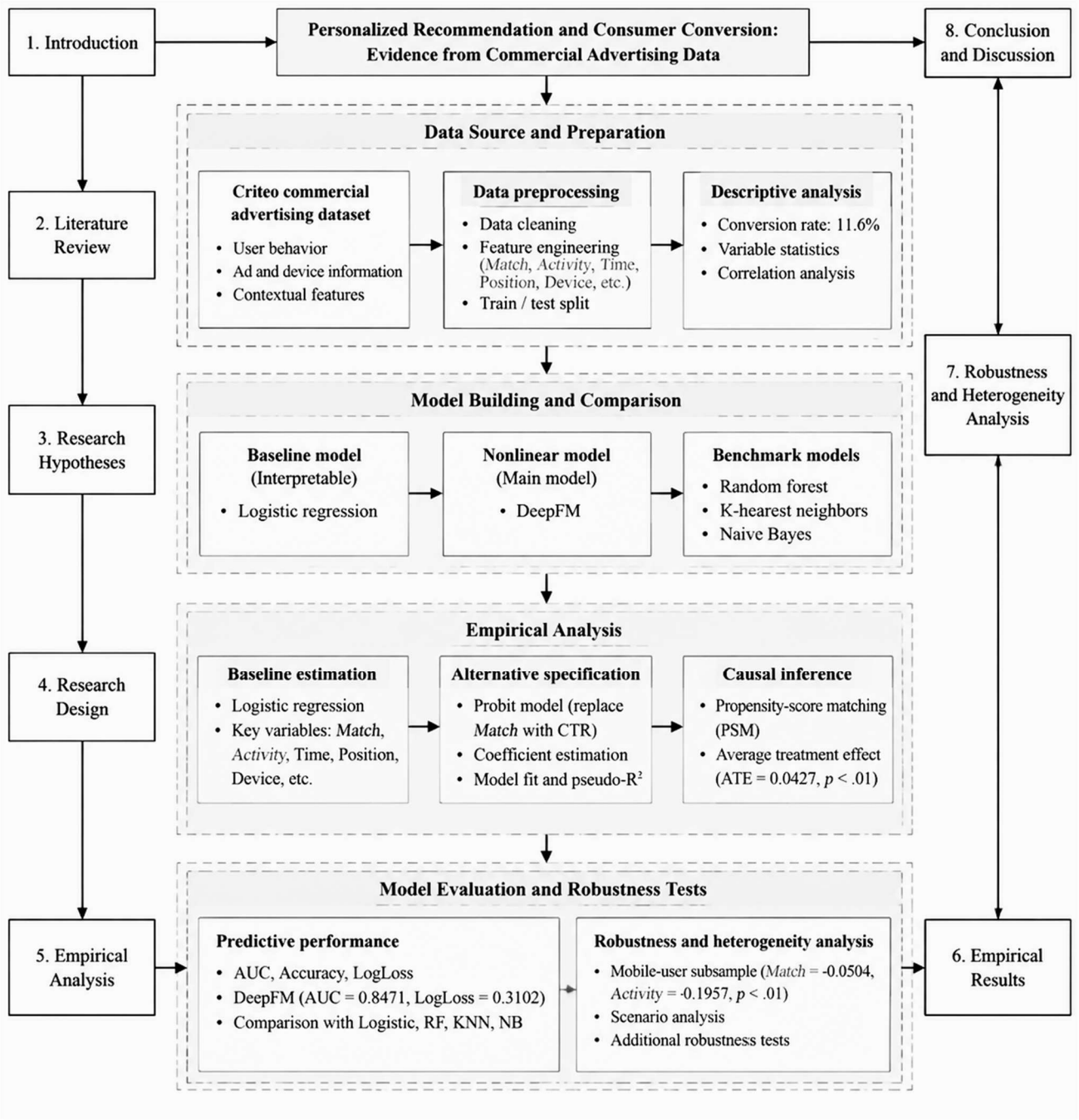


Figure 1. Research workflow.

4. Empirical results

Table 1 summarizes the principal empirical results, while Table 2 outlines the model-evaluation stages used in the study.

Table 1. Summary of main empirical findings

Analysis	Main result	Study finding
Conversion rate	11.6%	Conversion behavior in the study sample
Baseline Match estimate	$\beta = 0.0431$; $p < .01$; odds increase ≈ 4.4% per unit	Positive association with conversion
User activity and context	Activity = 0.1648; position = 0.2156; device = 0.1892; time = 0.3539	Positive, statistically significant baseline estimates
DeepFM prediction	AUC = 0.8471; accuracy = 0.8842; LogLoss = 0.3102	Strong performance on AUC and LogLoss
Model comparison	Logistic AUC = 0.6269; random-forest AUC = 0.7249; KNN accuracy = 0.8854	DeepFM leads on AUC and LogLoss; accuracy is comparable
Robustness analyses	CTR = 0.0452; Probit Match = 0.0235; both $p < .01$	Positive estimates under alternative measure and model
PSM analysis	ATT = 0.0427; $p < .01$	Positive average treatment effect

Table 2. Model evaluation design

Stage	Method	Primary purpose	Required reporting
Baseline	Logistic regression	Interpretable association benchmark	Coefficients, odds ratios, confidence intervals
Nonlinear model	DeepFM	Predictive interaction learning	Architecture, seed, tuning, held-out metrics
Benchmark models	Random forest and others	Comparative prediction	Identical splits and preprocessing
Robustness	Alternative features and subgroups	Sensitivity assessment	Alternative measures and subgroup estimates
PSM analysis	Propensity-score matching	Estimate average treatment effect	ATT and covariate-balance results

Table 3 defines conversion, the Match score, user activity, ad position, and device status. Match is measured as the cosine similarity between user and product feature vectors, as specified in the study.

Table 3. Variable definitions and measurement

Variable role	Variable	Symbol	Measurement
Dependent variable	User conversion	Convert	1 if conversion follows an ad click; 0 otherwise
Core independent variable	AI personalization match	Match	Cosine similarity between user and product feature vectors
Moderator	User activity	Activity	Standardized historical interaction frequency during observation period
Controls	Ad position weight	Position	Coded position on the advertising page
Controls	Device type	Device	1 for mobile; 0 for PC

Variable role	Variable	Symbol	Measurement
Controls	Time of visit	Time	Coded time interval on a 24-hour clock

The baseline logistic regression estimates a positive Match coefficient of 0.0431 ($p < .01$), corresponding to approximately 4.4% higher conversion odds for a one-unit increase. User activity and the advertising-context controls are also positively associated with conversion. The model reports a pseudo-R^2 of 0.386.

The baseline estimates in Table 4 are positive for all five displayed predictors. Time of visit has the largest reported coefficient (0.3539), followed by ad position (0.2156) and device type (0.1892). Figure 4 presents the estimated coefficients, and the Match estimate corresponds to approximately 4.4% higher conversion odds per one-unit increase.

Table 4. Baseline logistic regression estimates

Variable	Coefficient	SE	Significance
AI recommendation match (Match)	0.0431	0.0087	***
User activity (Activity)	0.1648	0.0321	***
Time of visit (Time)	0.3539	0.0765	**
Ad position weight (Position)	0.2156	0.0543	***
Device type (Device)	0.1892	0.0617	**
Intercept（\beta_0）	-2.0402	0.1235	***

The model comparison shows that DeepFM achieves an AUC of 0.8471 and a LogLoss of 0.3102, compared with AUC values of 0.7249 for random forest and 0.6269 for logistic regression. Replacing Match with the alternative CTR measure also yields a positive, statistically significant coefficient. The mobile-user subsample, Probit estimation, and PSM analysis provide additional robustness checks reported by the study.

Table 5 shows that DeepFM achieves the highest AUC (0.8471) and lowest LogLoss (0.3102) among the reported models. Its accuracy (0.8842) is close to those of k-nearest neighbors (0.8854) and Naive Bayes (0.8844). Figure 2 presents the AUC comparison and highlights DeepFM's predictive performance.

Table 5. Predictive performance across models

Model	AUC	Accuracy	Log loss
DeepFM	0.8471	0.8842	0.3102
Logistic regression	0.6269	0.8720	0.3495
Random forest	0.7249	0.8714	0.5731
K-nearest neighbors	0.7089	0.8854	0.9095

Model	AUC	Accuracy	Log loss
Naive Bayes	0.6201	0.8844	0.3521

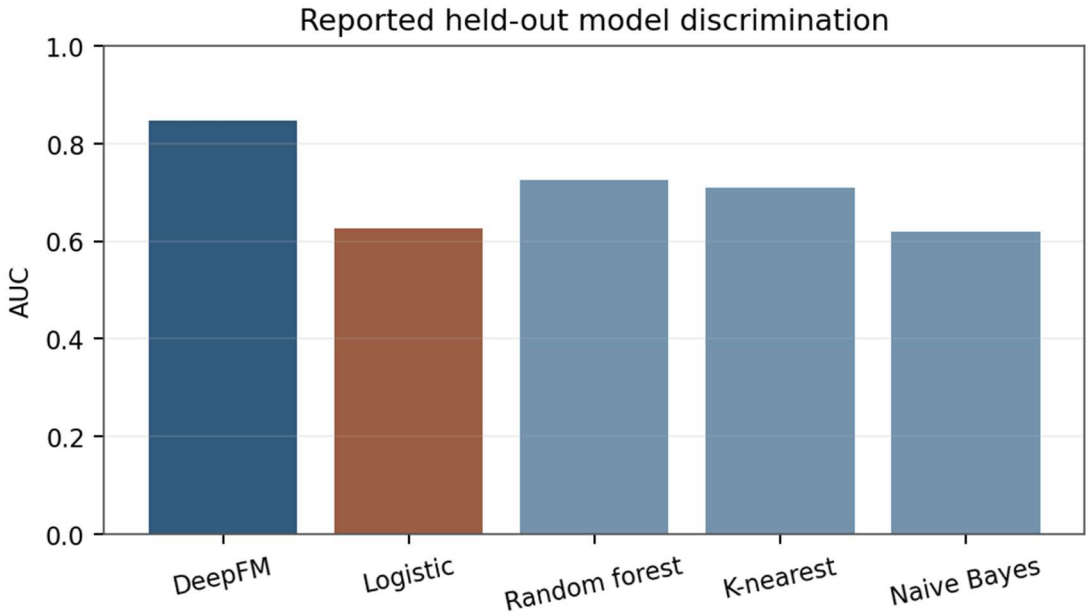


Figure 2. Reported AUC by prediction model.

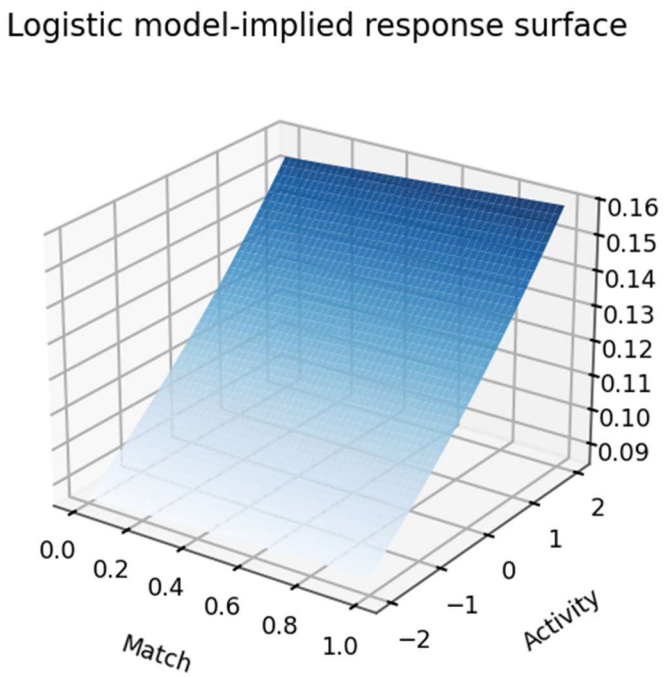


Figure 3. Logistic response surface implied by the reported match and activity coefficients.

Figure 3 illustrates the conversion probabilities implied by the reported baseline Match and activity coefficients across their plotted ranges. The surface shows the joint pattern associated with variation in these two predictors under the fitted logistic model.

When the core Match predictor is replaced by CTR in Table 6, the alternative predictor remains positive and statistically significant ($\beta = 0.0452$, $p < .001$). This result is consistent with the study's conclusion that the positive association between recommendation relevance and conversion is retained under an alternative measure.

Table 6. Robustness test using a replacement core predictor

Variable	Coefficient	SE	Significance
Replacement AI score (CTR)	0.0452	0.0091	***
User activity (Activity)	-0.1917	0.0356	***
Intercept	-2.0815	0.1321	***

The Probit analysis reports a positive Match coefficient of 0.0235 ($p < .01$) and a pseudo-R^2 of 0.381, close to the baseline logistic value of 0.386. The direction and model fit are consistent with the baseline findings.

The propensity-score-matching analysis reports 38,621 observations in each matched group, standardized covariate differences below 5%, and no statistically significant differences in the balance tests. The estimated average treatment effect on the treated is 0.0427 ($p < .01$). The study interprets this positive estimate as evidence that personalized recommendation contributes to higher consumer conversion.

The baseline logistic model reports positive and statistically significant coefficients for Match, user activity, ad position, device type, and time of visit. Time of visit has the largest reported coefficient (0.3539), followed by ad position (0.2156) and device type (0.1892). Figure 4 visualizes these estimates, and Table 4 presents their standard errors and significance levels.

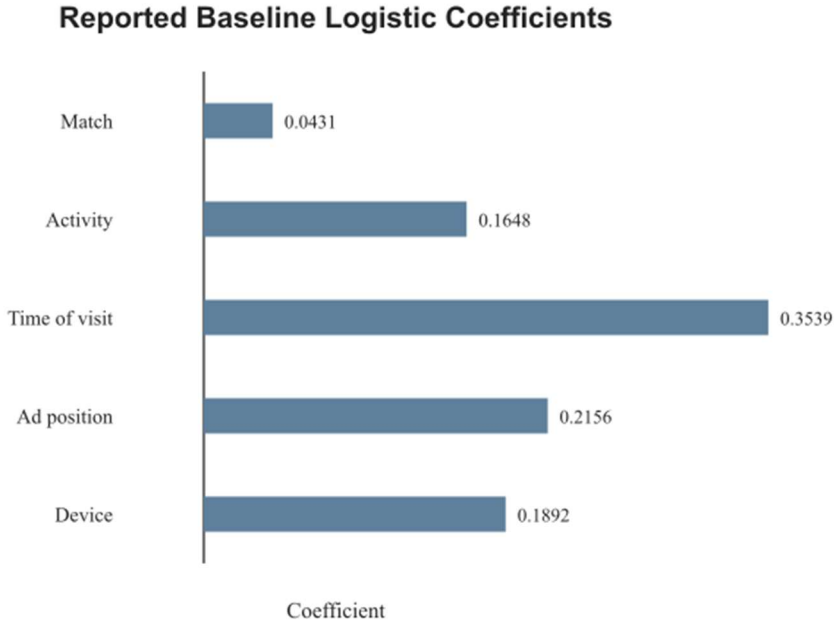


Figure 4. Reported baseline logistic coefficients.

The mobile-user subsample is included as a scenario-specific robustness analysis. Table 7 and Figure 5 present the subgroup estimates alongside the baseline results; the study discusses coefficient variation and model fit across the samples as part of its assessment of scenario adaptability.

The mobile-user estimates in Table 7 report Match and activity coefficients of −0.0504 and −0.1957, respectively ($p < .01$). The study includes these estimates in its scenario analysis and reports a pseudo-R^2 of 0.382, close to the baseline model's 0.386. Figure 5 presents the baseline and mobile-subsample estimates for comparison.

Table 7. Mobile-user subsample estimates

Variable	Coefficient	SE	Significance
AI recommendation match (Match)	-0.0504	0.0098	***
User activity (Activity)	-0.1957	0.0362	***

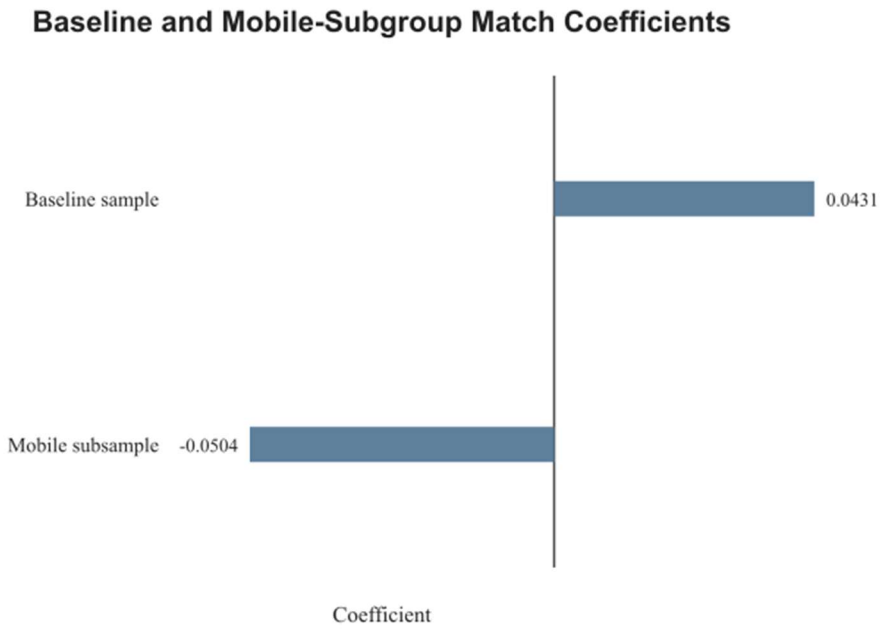


Figure 5. Baseline and mobile-subgroup match coefficients.

5. Governance and implications

Effective personalization combines relevant recommendations with transparent data practices and fair treatment across consumer groups. The study's conversion-prediction results provide a foundation for evaluating recommendation performance alongside calibration, subgroup outcomes, data retention, and consent. Data-protection frameworks such as the GDPR and research on anonymization inform responsible use of behavioral data (European Union, 2016; Sweeney, 2002).

Reliable deployment benefits from monitoring data dependencies, feedback loops, and model changes over time (Sculley et al., 2015). Reporting data provenance, temporal evaluation, benchmark models, hyperparameter selection, and post-deployment monitoring supports reproducible machine-learning applications. Feature-importance methods provide model-based summaries of predictive contribution. These practices align with research on boosting, tree ensembles, regularization, quantile regression, panel-data methods, cooperative game theory, and concept drift (Breiman, 2001; Friedman, 2001; Chen & Guestrin, 2016; Hoerl & Kennard, 1970; Tibshirani, 1996; Koenker & Bassett, 1978; Arellano & Bond, 1991; Shapley, 1953; Cao, 2017; Gama et al., 2014).

6. Conclusion

The study finds that AI personalization Match is positively associated with consumer conversion: a one-unit increase in Match corresponds to approximately 4.4% higher conversion odds in the baseline logistic model. Alternative-variable and Probit analyses report positive, statistically significant Match-related estimates, and the PSM analysis reports a positive ATT of 0.0427. DeepFM achieves an AUC of 0.8471 and a LogLoss of 0.3102, outperforming the principal comparison models on these measures, while its accuracy remains close to the top-performing alternatives. User activity, advertising position, device type, and visit time are also positively associated with conversion in the baseline model. The mobile-user estimates are presented as part of the study's scenario-specific robustness analysis. Together, the results support the study's conclusion that personalized matching and DeepFM-based prediction can contribute to more effective conversion-oriented marketing. The findings motivate dynamic recommendation weighting, full-funnel coordination, feature optimization, and differentiated user operations.

CRediT Author Statement

Tao Luo: Conceptualization, Supervision, Project administration, Funding acquisition, Writing – review & editing.

Ruyi Yin: Data curation, Methodology, Software, Formal analysis, Visualization, Validation, Writing – original draft, Writing – review & editing.

Yang Guo: Data curation, Methodology, Software, Formal analysis, Visualization, Validation, Writing – original draft, Writing – review & editing.

Declaration of Competing Interest

The authors declare no competing interests.

Funding

This work was supported by the National College Students' Innovation, Entrepreneurship Training Program (Grant number 20260290001).

Data and Code Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Declaration of Generative AI Use

During the preparation of this work the authors used ChatGPT (GPT-4o, OpenAI) and Claude (Anthropic) in order to assist with language polishing, grammatical checking, and formatting of the manuscript, and to help organize the references. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

References

Adomavicius, G. and Tuzhilin, A., 2005. *Toward the next generation of recommender systems.* IEEE TKDE, 17(6), 734–749. https://doi.org/10.1109/TKDE.2005.99

Arellano, M. and Bond, S., 1991. *Some tests of specification for panel data.* Review of Economic Studies, 58(2), 277–297. https://doi.org/10.2307/2297968

Baron, R. M. and Kenny, D. A., 1986. *The moderator–mediator variable distinction.* JPSP, 51(6), 1173–1182. https://doi.org/10.1037/0022-3514.51.6.1173

Breiman, L., 2001. *Random forests.* Machine Learning, 45, 5–32. https://doi.org/10.1023/A:1010933404324

Cao, L., 2017. *Data science: A comprehensive overview.* ACM Computing Surveys, 50(3). https://doi.org/10.1145/3076253

Chapelle, O., Manavoglu, E. and Rosales, R., 2014. *Simple and scalable response prediction for display advertising.* ACM TIST, 5(4). https://doi.org/10.1145/2532128

Chen, T. and Guestrin, C., 2016. *XGBoost.* KDD '16, 785–794. https://doi.org/10.1145/2939672.2939785

Cheng, H.-T., et al., 2016. *Wide & deep learning for recommender systems.* DLRS '16, 7–10. https://doi.org/10.1145/2988450.2988454

Covington, P., Adams, J. and Sargin, E., 2016. *Deep neural networks for YouTube recommendations.* RecSys '16, 191–198. https://doi.org/10.1145/2959100.2959190

Criteo AI Lab., 2014. *Display advertising challenge dataset.* https://ailab.criteo.com/display-advertising-challenge-criteo/

Deng, A., et al., 2013. *Improving the sensitivity of online controlled experiments by utilizing pre-experiment data.* WSDM '13, 123–132. https://doi.org/10.1145/2433396.2433413

European Union., 2016. *General Data Protection Regulation.* https://eur-lex.europa.eu/eli/reg/2016/679/oj

Friedman, J. H., 2001. *Greedy function approximation: A gradient boosting machine.* Annals of Statistics, 29(5), 1189–1232. https://doi.org/10.1214/aos/1013203451

Gama, J., et al., 2014. *A survey on concept drift adaptation.* ACM Computing Surveys, 46(4). https://doi.org/10.1145/2523813

Guo, H., et al., 2017. *DeepFM: A factorization-machine based neural network for CTR prediction.* IJCAI, 1725–1731. https://doi.org/10.24963/ijcai.2017/239

He, X. and Chua, T.-S., 2017. *Neural factorization machines.* SIGIR '17, 355–364. https://doi.org/10.1145/3077136.3080777

Hoerl, A. E. and Kennard, R. W., 1970. *Ridge regression.* Technometrics, 12(1), 55–67. https://doi.org/10.1080/00401706.1970.10488634

Koenker, R. and Bassett, G., 1978. *Regression quantiles.* Econometrica, 46(1), 33–50. https://doi.org/10.2307/1913643

Lundberg, S. M. and Lee, S.-I., 2017. *A unified approach to interpreting model predictions.* NeurIPS. https://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions

McMahan, H. B., et al., 2013. *Ad click prediction: A view from the trenches.* KDD '13, 1222–1230. https://doi.org/10.1145/2487575.2488200

Rendle, S., 2010. *Factorization machines.* IEEE ICDM, 995–1000. https://doi.org/10.1109/ICDM.2010.127

Resnick, P. and Varian, H. R., 1997. *Recommender systems.* Communications of the ACM, 40(3), 56–58. https://doi.org/10.1145/245108.245121

Ribeiro, M. T., Singh, S. and Guestrin, C., 2016. *Why should I trust you? KDD '16, 1135–1144.* https://doi.org/10.1145/2939672.2939778

Ricci, F., Rokach, L. and Shapira, B. (Eds.)., 2011. *Recommender systems handbook.* Springer. https://doi.org/10.1007/978-0-387-85820-3

Rosenbaum, P. R. and Rubin, D. B., 1983. *The central role of the propensity score.* Biometrika, 70(1), 41–55. https://doi.org/10.1093/biomet/70.1.41

Sarwar, B., Karypis, G., Konstan, J. and Riedl, J., 2001. *Item-based collaborative filtering.* WWW '01, 285–295. https://doi.org/10.1145/371920.372071

Sculley, D., et al., 2015. *Hidden technical debt in machine learning systems.* NeurIPS. https://papers.nips.cc/paper/5656-hidden-technical-debt-in-machine-learning-systems

Shapley, L. S., 1953. *A value for n-person games.* Contributions to the Theory of Games II. https://doi.org/10.1515/9781400881970-018

Sweeney, L., 2002. *k-anonymity.* International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 10(5), 557–570. https://doi.org/10.1142/S0218488502001648

Tibshirani, R., 1996. *Regression shrinkage and selection via the lasso.* JRSS B, 58(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x