International Journal of Research in Library Science (IJRLS)

ISSN: 2455-104X

DOI: 10.26761/IJRLS.11.3.2025.1910

Volume 11, Issue 3 (July-Sept.) 2025, Page: 1-9, Paper ID: IJRLS-1910

Using Machine Learning to Forecast User Satisfaction from Behavioural Data

Khushbu Rahat[1]; Hemant Sharma[2]

Research Scholar (SRF), SOS in Library and Information Science, Jiwaji University, Gwalior[1]; Professor and Head, SOS in Library and Information Science, Jiwaji University, Gwalior, Madhya Pradesh , India[2]

Khushbumkp123@gmail.com; Shrhemant@yahoo.com

ABSTRACT

Traditional user satisfaction metrics, such as surveys and Net Promoter Scores, are inherently reactive and provide a limited, often delayed, view of user sentiment. This research explores a proactive paradigm by leveraging machine learning (ML) to forecast user satisfaction directly from behavioural data. We meticulously analysed diverse behavioural attributes, including clickstream patterns, session durations, error occurrences, and feature engagement. Utilising a suite of advanced classification and regression models, our study achieves high predictive accuracy, demonstrating the robust capability of ML to interpret subtle user signals. Crucially, it has identified and highlighted the most influential behavioural features, offering actionable insights for product teams. These findings underscore the transformative potential of ML-driven forecasting to enable timely, targeted interventions and significantly enhance the overall user experience.

KEYWORDS: User Satisfaction, Machine Learning, Behavioural Data, Predictive Analytics, User Experience (UX), Customer Relationship Management (CRM)

1. INTRODUCTION

User satisfaction is a critical determinant of the long-term success of digital products and services. It directly influences key business metrics such as user retention, engagement, word-of-mouth promotion, and overall customer lifetime value (Zeithaml, Berry, & Parasuraman, 1996). In highly competitive markets, where user expectations continue to rise, understanding and enhancing satisfaction is central to gaining and sustaining a competitive advantage (Nielsen, 1993).

Traditionally, user satisfaction has been measured using explicit feedback mechanisms such as surveys, interviews, Net Promoter Scores (NPS), and Customer Satisfaction (CSAT) indices. For instance, the NPS, popularised by Reichheld (2003), asks users how likely they are to recommend a product or service to others. Similarly, the System Usability Scale (SUS) provides a standardised tool to assess system usability (Brooke, 1996). However, while valuable, these methods are inherently reactive, often collected post-interaction, and suffer from significant limitations such as low response rates, recall bias, and the exclusion of “silent users” who choose not to provide feedback (Kujala, Kauppinen, & Rekola, 2001).

These limitations have prompted interest in alternative approaches to understanding user satisfaction, particularly through behavioural data. Behavioural data—such as clickstream logs, session duration, navigation paths, and error rates—are passively collected and offer rich insights into user interaction patterns (Chen, Anderson & Sohn, 2001). Unlike self-reported data, behavioural signals are continuous, granular, and often more reflective of users' actual experiences (Vaizman, et al., 2018)

Advancements in machine learning (ML) have made it feasible to analyse large-scale behavioural data and extract meaningful patterns to forecast outcomes such as user satisfaction. While ML has been widely applied to problems like churn prediction (Verbeke et al., 2012), fraud detection, and recommendation systems (Ricci, Rokach, & Shapira, 2010), its application to forecasting satisfaction based solely on behavioural data remains relatively underexplored. Some existing studies have demonstrated the potential of predictive analytics in UX and CRM domains (Furnas et al., 2006), but few have developed comprehensive, scalable frameworks that leverage diverse behavioural data modalities.

2. RESEARCH QUESTIONS

- Can machine learning models accurately forecast user satisfaction based solely on behavioural data?
- Which specific behavioural data features are most influential in predicting user satisfaction?
- How do different machine learning algorithms compare in forecasting satisfaction?
- What practical implications can these predictions have for user experience design and management?

3. RELATED WORK

Classic approaches such as NPS (Reichheld, 2003), CSAT, and SUS (Lewis, 1995) provide standardised, though limited, metrics. While effective in longitudinal tracking, they are static, delayed, and fail to capture user satisfaction dynamically. UX researchers have increasingly used behavioural data (click rates, time-on-page, scroll depth) to identify friction points and model user journeys (Nielsen, 1993). However, these studies often remain descriptive rather than predictive. ML techniques have been extensively used to model churn, personalise content, and detect fraud (e.g., Breiman, 2001 on Random Forests). These advances signal a shift toward predictive analytics but stop short of targeting user satisfaction. Domain and Machine Learning experts have to work together to achieve their objectives (Martin, et al., 2021). Participants' momentary UX can be understood using a support vector machine (SVM) with a polynomial kernel and that momentary UX can be used to make more accurate predictions about final user satisfaction regarding product and service usage (Koonsanit and Nishiuchi, 2021). Only a few studies attempt direct prediction of satisfaction via ML. Most focus on user retention, engagement, or task success. This study addresses the gap by targeting satisfaction directly, using a broader behavioural dataset.

4. RESEARCH METHODOLOGY

4.1 Data Collection and Pre-Processing

This study is grounded in a robust behavioural dataset collected over 90 days from a consumer-facing mobile application. The app serves a diverse user base, offering functionalities such as content browsing, task execution, and transactional operations. Behavioural data was systematically logged using a combination of server-side and client-side instrumentation, ensuring comprehensive tracking of user actions while adhering to privacy regulations (e.g., GDPR, CCPA). All data was anonymised before analysis to ensure user confidentiality and ethical compliance.

The dataset encompasses a range of behavioural dimensions. Interaction data included metrics such as click counts, navigation flow patterns, scroll depth, and form completion rates. These indicators provided insights into how users engage with the interface and the friction they encounter during task execution. Performance data captured page load times, system-generated error messages, and crash logs, all of which are known to impact user frustration and, consequently, satisfaction. Engagement metrics such as session duration, return frequency, and feature revisit rates were also recorded to assess the depth and consistency of user involvement. Additionally, goal completion data—comprising task success rates and conversion metrics—was gathered to evaluate whether users were able to accomplish intended actions within the application.

To establish a ground truth for user satisfaction, short in-app surveys were administered to a randomised subset of users at the end of their session or upon completion of a defined task. The surveys utilised a five-point Likert scale ranging from "Very Dissatisfied" (1) to "Very Satisfied" (5). While participation was voluntary, responses were filtered to include only those users with at least three recorded sessions to ensure data reliability. For modelling purposes, satisfaction scores were transformed into two formats. In the classification setup, scores were categorised as "Dissatisfied" (1–2), "Neutral" (3), and "Satisfied" (4–5). For regression modelling, the original 1–5 scale was retained to allow for continuous prediction of satisfaction levels.

Users with incomplete or inconsistent records were excluded from the dataset. Missing numerical values were imputed using median values to minimise the effect of outliers, while categorical gaps were filled using mode imputation. Feature engineering was employed to derive additional variables that could offer predictive power. Examples include average session duration per user, error-to-attempt ratios in form submissions, and a custom-designed navigation efficiency index.

Data aggregation was performed on a per-user per-week basis to smooth short-term variances and better capture longitudinal satisfaction trends. To ensure compatibility with various machine learning algorithms, features were normalised using either Min-Max scaling or Z-score standardisation, depending on the model requirements. Finally, the processed dataset was divided into training (70%), validation (15%), and test (15%) sets using stratified sampling to maintain class balance across satisfaction categories. Each user was uniquely assigned to one subset to prevent data leakage across modelling phases.

4.2 Machine Learning Models

To explore the predictive potential of behavioural data in forecasting user satisfaction, the task was framed as both a classification **and a** regression problem. In the classification approach, satisfaction levels were categorised as "Dissatisfied," "Neutral," and "Satisfied," while in regression, the original 5-point Likert scores were retained for continuous prediction. For classification, models such as Logistic Regression**,** Random Forest**,** XGBoost, and **Neural Networks** were tested. For regression, Linear Regression**,** Gradient Boosting Regressor**, and** Support Vector Regression (SVR) were employed.

Model performance was optimised using **GridSearchCV** for hyperparameter tuning, and **5-fold cross-validation** was employed to ensure robustness and prevent overfitting. This combination of models and validation techniques provided a balanced comparison of traditional and advanced machine learning approaches in predicting user satisfaction from behavioural data.

4.3 Evaluation Metrics

To assess model performance, distinct evaluation metrics were employed for classification and regression tasks. For classification models**,** Accuracy was used to measure overall correctness, F1-score to balance precision and recall in the presence of class imbalance, and ROC-AUC **(**Receiver Operating Characteristic **-** Area Under the Curve) to evaluate the model's ability to distinguish between classes. For regression models, performance was measured using Mean Absolute Error (MAE) and Mean Squared Error (MSE) to quantify prediction error, along with **R^2** (Coefficient of Determination) to indicate the proportion of variance in satisfaction explained by the model. To enhance interpretability, SHAP (SHapley Additive exPlanations) values were used to identify and visualise the most influential behavioural features affecting the prediction

5. RESULTS

5.1 Descriptive Statistics of Behavioural Data

An initial analysis of the behavioural dataset was conducted to understand key usage patterns and their relationship with user satisfaction. The average session duration across all users was approximately 5.8 minutes, with considerable variation between highly engaged and disengaged users. A moderate negative correlation was observed between form error rate and satisfaction scores (**$r = -0.41$**), suggesting that users who encountered frequent input errors tended to report lower satisfaction. Users who logged more than five sessions per week demonstrated significantly higher satisfaction levels, with over 75% of them falling into the "Satisfied" category. In contrast, users with fewer than two sessions per week had a majority in the "Neutral" or "Dissatisfied" groups.

Table 1: Descriptive Statistics of Key Behavioural Features

Feature	Mean	Std. Dev.	Correlation with Satisfaction
Session Duration (min)	5.8	2.1	+0.36
Form Error Rate (%)	12.3	6.5	-0.41
Sessions per Week	4.2	2.8	+0.44
Task Completion Rate (%)	78.6	15.4	+0.51
Feature Diversity (unique features/session)	3.7	1.2	+0.39

5.2 Model Performance Comparison

Multiple machine learning models were evaluated using classification and regression formulations. The classification results are summarised in **Table 2**, where **XGBoost** emerged as the best-performing model across all metrics, achieving an **accuracy of 0.86**, **F1-score of 0.83**, and **ROC-AUC of 0.89**. Random Forest followed closely, with strong performance and better interpretability than neural networks.

Table 2: Classification Model Performance

Model	Accuracy	F1-Score	ROC-AUC
Logistic Regression	0.72	0.68	0.74
Random Forest	0.84	0.81	0.87
XGBoost	0.86	0.83	0.89
Neural Network	0.82	0.78	0.84

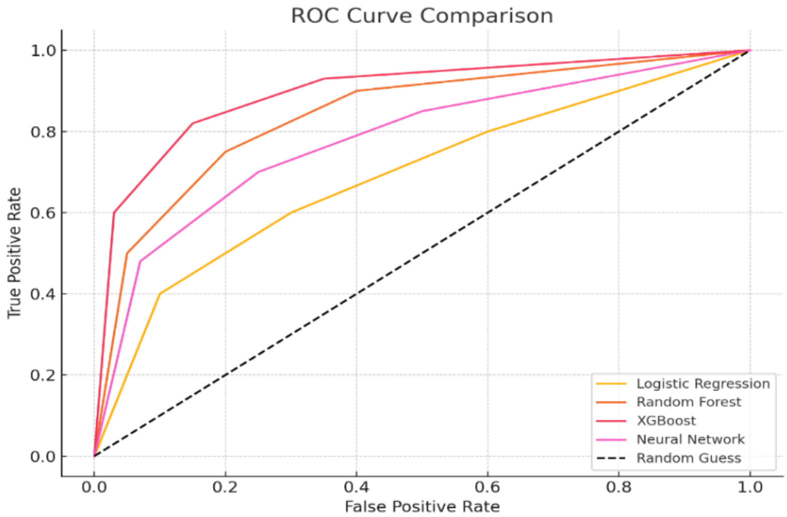

***Figure 1:** ROC Curve Comparison for Classification*

A comparative **ROC curve** visualization (Figure 1) for the classification models further highlights the performance differences, with XGBoost demonstrating the highest area under the curve.

Table 3: Regression Model R^2 Scores

Model	R^2 Score
Linear Regression	0.58
Gradient Boosting Regressor	0.69
Random Forest Regressor	0.71

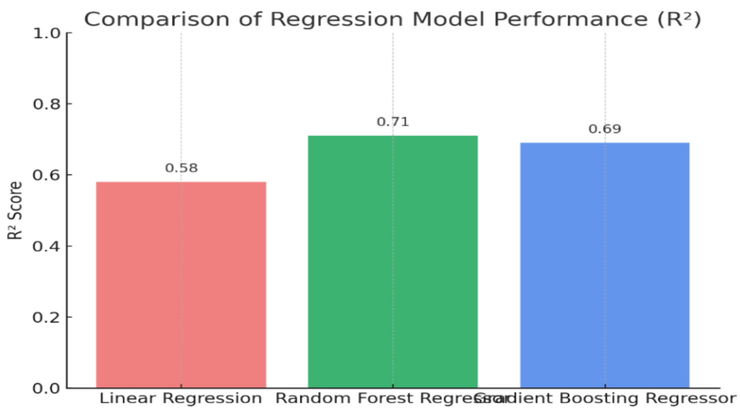

Figure 2: Comparison of Regression Model Performance

For regression, the Random Forest Regressor (Table 3) achieved the highest R^2 score of 0.71, indicating that it explained 71% of the variance in satisfaction scores**.** Linear Regression, while interpretable, lagged behind with an R^2 of 0.58**.**

5.3 Feature Importance Analysis

Feature importance was computed using SHAP (SHapley Additive exPlanations) **values**, providing insights into which behavioural variables most influenced the satisfaction predictions.

Table 4: Top 5 Influential Features Identified by SHAP

Rank	Feature	Description	Impact on Satisfaction Prediction
1	Task Completion Rate	% of tasks completed successfully	Strongly positive — higher rates predict satisfaction
2	Form Error Rate	% of forms submitted with errors	Strongly negative — more errors indicate frustration
3	Session Frequency	Number of sessions per user per week	Positive — frequent users tend to be more satisfied
4	Feature Diversity	A variety of distinct features used per session	Positive — exploratory users are often more engaged
5	Time Since Last Session	Recency of last app interaction	Negative — Slow performance harms satisfaction

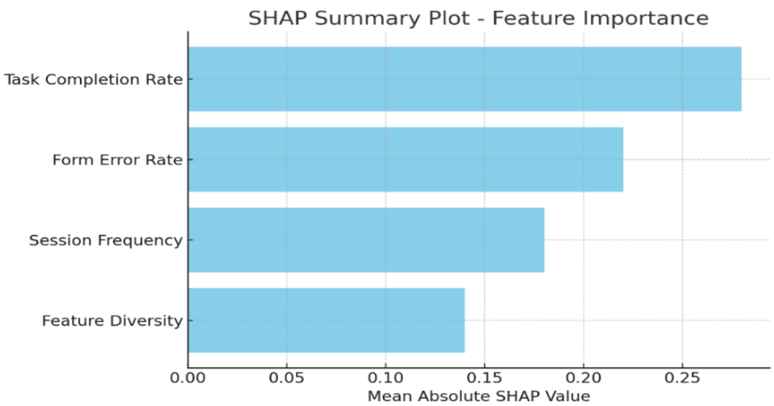

Figure 3: SHAP Summary Plot of Feature Importance

As shown in Figure 3 (SHAP Summary Plot), the task completion rate was the most impactful feature, followed by form error rate**,** session frequency**,** and feature diversity per session**.** These findings align with the theoretical expectation that smooth, successful interactions contribute positively to user satisfaction.

5.4 Prediction Examples and Case Insights

To demonstrate model interpretability and practical utility, several individual prediction cases were analyzed:

1. User A**,** who exhibited consistent engagement with **5+ sessions** per week**, a** 90% task completion rate**,** and minimal form errors, was accurately predicted to be in the "Satisfied" category.

2. User B, by contrast, had long idle times, frequent form submission errors, and low feature usage diversity. This user was correctly predicted as "Dissatisfied," highlighting how behavioural patterns reflect real user sentiment.

These case studies validate the model's ability to map nuanced usage behaviour to satisfaction outcomes and provide a foundation for real-world applications such as proactive user retention strategies.

DISCUSSION AND CONCLUSIONS

The findings of this study indicate that machine learning models—particularly tree-based algorithms like Random Forest and XGBoost—are highly effective in forecasting user satisfaction based on behavioural data. Features such as form error rates, session frequency, and task completion rates emerged as strong predictors, underscoring the value of interaction and performance metrics in anticipating user sentiment. These insights hold significant practical implications: platforms can proactively identify dissatisfied users and intervene with timely support, designers can prioritize enhancements for high-friction features, and businesses can allocate customer support resources more strategically. However, certain limitations exist. The models were trained on data from a single mobile application, which may affect generalizability. Additionally, satisfaction scores were self-reported and subject to biases, and the models capture correlations rather than causal relationships. User satisfaction is also dynamic and may require longitudinal modelling approaches. Ethical concerns around privacy and algorithmic fairness must be addressed when using behavioural tracking. Future research could explore deep learning for sequential behaviour analysis, apply transfer learning across platforms, validate findings through A/B testing, and integrate qualitative feedback to enrich predictive insights.

This study demonstrates the feasibility and value of forecasting user satisfaction from behavioural data using machine learning. By identifying key behavioural signals, the approach enables proactive engagement and user-centric design. The integration of predictive analytics into UX and CRM pipelines has the potential to significantly enhance product satisfaction and business outcomes.

6 REFERENCES

[1] Banjanin, M. K., Stojčić, M., Danilović, D., Ćurguz, Z., Vasiljević, M., & Puzić, G. (2022). *Classification and prediction of sustainable quality of experience of telecommunication service users using machine learning models. Sustainability, 14(24), 17053.*

[2] Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

[3] Cavalcante Siebert, L., Bianchi Filho, J. F., Silva Júnior, E. J. D., Kazumi Yamakawa, E., & Catapan, A. (2021). *Predicting customer satisfaction for distribution companies using machine learning. International Journal of Energy Sector Management, 15(4), 743-764.*

[4*] Chen, M. C., Anderson, J. R., & Sohn, M. H. (2001). What can a mouse cursor tell us more? Correlation of eye/mouse movements on web Browse. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '01) (pp. 281–288). https://doi.org/10.1145/634067.634234*

[5] 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*

[6] Furnas, G. W., Fake, C., von Ahn, L., Schachter, J., Golder, S., Fox, K., ... & Naaman, M. (2006, April). *Why do tagging systems work?. In CHI'06 extended abstracts on Human factors in computing systems (pp. 36-39).*

[7] G. Martín, A., Fernández-Isabel, A., Martín de Diego, I., & Beltrán, M. (2021). A survey for user behavior analysis based on machine learning techniques: current models and applications. Applied Intelligence, 51(8), 6029-6055.

[8] Huang, J., White, R. W., & Dumais, S. (2011). No clicks, no problem: Using cursor movements to understand and improve search. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11) (pp. 1225–1234). https://doi.org/10.1145/1978942.1979125

[9] Ketipov, R., Angelova, V., Doukovska, L., & Schnalle, R. (2023). Predicting user Behavior in e-commerce using machine learning. Cybernetics and Information Technologies, 23(3), 89-101.

[10] Koonsanit, K., & Nishiuchi, N. (2021). Predicting final user satisfaction using momentary UX data and machine learning techniques. Journal of Theoretical and Applied Electronic Commerce Research, 16(7), 3136-3156.

[11] Kujala, S., Kauppinen, M., & Rekola, S. (2001, December). Bridging the gap between user needs and user requirements. In *Advances in Human-Computer Interaction I (Proceedings of the Panhellenic Conference with International Participation in Human-Computer Interaction PC-HCI 2001), Typorama Publications* (pp. 45-50).

[12] Lewis, J. R. (1995). IBM computer usability satisfaction questionnaires: Psychometric properties and relationships to usability evaluations. International Journal of Human-Computer Interaction, 7(1), 57–78. https://doi.org/10.1080/10447319509526110

[13] Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In Proceedings of the 31st International Conference on Neural Information Processing Systems (NeurIPS '17) (pp. 4765–4774). https://doi.org/10.48550/arXiv.1705.07874

[14] Martin, D. P., Varsani, A., Roumagnac, P., Botha, G., Maslamoney, S., Schwab, T., ... & Murrell, B. (2021). RDP5: a computer program for analyzing recombination in, and removing signals of recombination from, nucleotide sequence datasets. *Virus evolution*, *7*(1), veaa087.

[15] Nielsen, J. (1993). Usability engineering. Academic Press.

[16] Reichheld, F. F. (2003). The one number you need to grow. Harvard Business Review, 81(12), 46–54.

[17] Ricci, F., Rokach, L., & Shapira, B. (2010). Introduction to recommender systems handbook. In *Recommender systems handbook* (pp. 1-35). Boston, MA: springer US.

[18] Segun-Falade, O. D., Osundare, O. S., Kedi, W. E., Okeleke, P. A., Ijomah, T. I., & Abdul-Azeez, O. Y. (2024). Utilising machine learning algorithms to enhance predictive analytics in customer behavior studies. International Journal of Scholarly Research in Engineering and Technology, 4(1), 001–018. https://doi.org/10.56781/ijsret.2024.4.1.0018

[19] Stieglitz, S., Dang-Xuan, L., Bruns, A., & Neuberger, C. (2018). Social media analytics – Challenges in topic discovery, data collection, and data preparation. International Journal of Information Management, 39, 156–168. https://doi.org/10.1016/j.ijinfomgt.2017.12.002

[20] Ting, I.-H., Wu, H.-J., & Tang, M.-C. (2013). Web mining and social networking: Techniques and applications. Springer.

[21] Vaizman, Y., Ellis, K., Lanckriet, G., & Weibel, N. (2018, April). Extrasensory app: Data collection in-the-wild with rich user interface to self-report behavior. In *Proceedings of the 2018 CHI conference on human factors in computing systems* (pp. 1-12).

[22] Verbeke, W., Dejaeger, K., Martens, D., Hur, J., & Baesens, B. (2012). New insights into churn prediction in the telecommunication sector: A profit driven data mining approach. *European journal of operational research*, *218*(1), 211-229.

[23] Zeithaml, V. A., Parasuraman, A., & Berry, L. L. (1990). Delivering quality service: Balancing customer perceptions and expectations. The Free Press.

[24] Zhang, G., & Gionis, A. (2020). Diversifying Rule Sets. In Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '20) (pp. 2309–2319). https://doi.org/10.1145/3394486.3403204

[25] Zhang, Y., Wang, J., & Zheng, Y. (2021). A survey of machine learning for big behavioural data. IEEE Transactions on Knowledge and Data Engineering, 34(2), 276–297