VOLUME 01 | ISSUE 10
OCTOBER 2026

MULTIDISCIPLINARY • PEER-REVIEWED
OPEN ACCESS • PAGES 15–20

SRJ: International Journal of Scientific Research
ISSN(O): 3139-8758

ORIGINAL RESEARCH ARTICLE

**Article DOI:** 10.68295/SRJ/XPGP2552
**Article URL:** https://doi.org/10.68295/SRJ/XPGP2552

# AN INTELLIGENT CONSUMER BEHAVIOUR PREDICTION SYSTEM USING HYBRID MACHINE LEARNING ALGORITHMS

**[1]Ms. Munesh Yadav, [2]Mr. Dilip Kumar and [3]Mr. Praveen Kumar**

1. *Assistant Professor, School of Engineering & Technology, Starex University, Gurugram, India*
2. *Assistant Professor, School of Computer Science, Starex University, Gurugram, India*
3. *Assistant Professor, School of Engineering & Technology, Starex University, Gurugram, India*

**MANUSCRIPT INFO**

Received:18 Aug 2026
Revised: 19 Sep 2026
Final Accepted: 09 Oct 2026
Published: 22 Oct 2026

**KEYWORDS**
consumer behaviour prediction;
hybrid machine learning;
stacked ensemble; purchase intention;
e-commerce analytics;
K-Means clustering

**ABSTRACT**

Predicting whether a website visitor will make a purchase is one of the most commercially valuable problems in online marketing, yet it is difficult because conversion is rare and visitor behaviour is noisy. This paper presents an intelligent consumer behaviour prediction system that combines unsupervised segmentation with a stacked ensemble of Random Forest, XGBoost and LightGBM learners, joined by a logistic-regression meta-learner. The system was trained and evaluated on the public Online Shoppers Purchasing Intention dataset, which holds 12,330 browsing sessions of which 15.5% ended in a purchase. After cleaning, encoding and engineering three aggregate behavioural features, an 80/20 stratified split was used for evaluation. The hybrid model reached an AUC of 0.929, the highest of all nine configurations tested, and, after a decision threshold was tuned on training data only, an F1-score of 0.678 with a recall of 0.715 on the held-out sessions. A baseline ablation shows that most of the gain comes from stacking heterogeneous learners, while the clustering step mainly contributes interpretability by separating a segment with a 29.4% conversion rate from one with 1.8%. Page value, exit rate and time spent emerged as the strongest predictors. The paper discusses managerial uses of the system and the limits of a purely behavioural data source, and proposes the inclusion of psychological signals such as fear of missing out, urge for instant gratification and social validation in later versions.

**Corresponding Author:** ***Ms. Munesh Yadav***
*Assistant Professor, School of Engineering & Technology, Starex University, Gurugram, India*

© 2026 by the Author(s). Published by SRJ Publication and Research LLP under the Creative Commons Attribution (CC BY) 4.0 License.

## 1. Introduction

Online retailers observe millions of sessions, and the great majority of them end without a sale. In the dataset analysed here, roughly five of every six visits never convert. For a marketing team this raises a practical question: which visitors are worth intervening on, and when? An accurate, real-time estimate of purchase probability allows discounts, chat prompts and remarketing budgets to be targeted at the visitors who are undecided, rather than spread across everyone.

Consumer behaviour is shaped by more than clicks. Gaur, Sehgal, Sharma, Gaur and Kargeti (2026) argue that fear of missing out, the desire for instant gratification and the need for social validation all push consumers toward quicker online purchase decisions. Such psychological forces leave traces in browsing logs: short, focused sessions on high-value pages, repeated product views, and low exit rates are all plausible footprints of a motivated buyer. A prediction system that reads these traces can therefore act as an indirect sensor of intent, even when the underlying motives cannot be measured directly.

Single classifiers rarely capture every pattern in such data. Linear models miss interactions, margin-based models struggle with skewed classes, and individual trees are unstable. Hybrid designs, which combine different learning paradigms, are a common remedy. This study builds one such design and tests it honestly against its own components.

The contributions of the paper are as follows:

- A hybrid pipeline that merges unsupervised customer segmentation with a stacked ensemble of three complementary tree-based learners.
- A transparent comparison against six standard classifiers and an ablation that isolates the effect of each hybrid component.
- An interpretation of the model through feature importance and cluster profiles, linking results to managerial action.
- A reproducible protocol using only openly available data.

The rest of the paper is organised as follows. Section 2 reviews related work, Section 3 describes the data and method, Section 4 reports results, Section 5 discusses implications and limits, and Section 6 concludes.

## 2. Related Work

Research on predicting online purchases has moved from session-level statistics toward learning-based systems. Sakar, Polat, Katircioglu and Kastro (2019) built the dataset used in this study and showed that page-level metrics from web analytics, combined with visitor and calendar attributes, can drive real-time prediction of shopping intent with neural networks. Their work established that navigation behaviour carries strong purchase signals, and the data have since become a common benchmark.

Ensemble learning is a second relevant strand. Breiman (2001) showed that averaging many decorrelated decision trees reduces variance and improves generalisation. Gradient boosting was later scaled to large tabular problems by Chen and Guestrin (2016) in XGBoost and by Ke et al. (2017) in LightGBM, which uses histogram-based splitting and leaf-wise growth for speed. Wolpert (1992) proposed stacked generalisation, in which a meta-learner is trained on the out-of-fold predictions of base learners so that it can learn where each base model is reliable. Dietterich (2000) explained why combining diverse models can outperform the best single member.

A third strand is segmentation. Partitioning customers with K-Means (Lloyd, 1982) before supervised modelling has been used to expose latent groups with distinct behaviour, either as input features or as separate models per segment.

Finally, the behavioural-science literature offers explanations for why consumers act as they do online. Gaur et al. (2026) examine how fear of missing out, instant gratification and social validation influence consumer behaviour in online marketing. Their findings motivate the present system as a tool for detecting, in log data, the outcomes of those psychological pressures, and they identify a gap that this paper acknowledges: purely behavioural logs do not contain the psychographic variables themselves.

## 3. Methodology

### 3.1 System overview

The proposed system has five stages, shown in order of execution: (1) data acquisition and cleaning, (2) feature encoding and engineering, (3) unsupervised segmentation, (4) stacked ensemble learning, and (5) threshold calibration and evaluation. The segmentation output is passed to the ensemble as additional input, so the two paradigms are combined in a single model rather than run side by side.

### 3.2 Dataset

The Online Shoppers Purchasing Intention dataset, donated to the UCI Machine Learning Repository in 2018 and described by Sakar et al. (2019), was used. Each of its 12,330 rows is a separate session, and the sessions were collected so that each belongs to a different user over one year, which limits bias toward a specific campaign, special day or user profile. The target variable, Revenue, equals one when the session ended in a purchase. A total of 1,908 sessions (15.5%) are positive, which makes the problem moderately imbalanced. The file has no missing values. Table 1 summarises the variables.

**Table 1. Description of the dataset variables**

| Group | Variables | Type |
|---|---|---|
| Page counts | Administrative, Informational, ProductRelated | Integer |
| Time on page type (seconds) | Administrative_Duration, Informational_Duration, ProductRelated_Duration | Continuous |
| Analytics metrics | BounceRates, ExitRates, PageValues | Continuous |
| Calendar | SpecialDay, Month, Weekend | Numeric, categorical, binary |
| Technical context | OperatingSystems, Browser, Region, TrafficType | Integer codes |
| Visitor profile | VisitorType (new, returning, other) | Categorical |
| Target | Revenue (purchase = 1) | Binary |

Source: Sakar et al. (2019); UCI Machine Learning Repository.
Returning visitors made up 10,551 of the sessions, new visitors 1,694 and other visitors 85. New visitors converted at 24.9%, against 13.9% for returning visitors. Weekend sessions converted at 17.4%, against 14.9% on weekdays. These descriptive contrasts are summarised in Table 2.

**Table 2. Conversion rate by visitor and day type (n = 12,330)**

| Segment | Sessions | Purchase rate |
|---|---|---|
| New visitor | 1,694 | 24.9% |
| Returning visitor | 10,551 | 13.9% |
| Other | 85 | 18.8% |
| Weekday | 9,462 | 14.9% |
| Weekend | 2,868 | 17.4% |
| All sessions | 12,330 | 15.5% |

### 3.3 Preprocessing and feature engineering

Month and VisitorType were one-hot encoded, with the first level dropped, and the Weekend and Revenue flags were converted to binary values. Three aggregate features were engineered to capture overall engagement: TotalPages (the sum of the three page counts), TotalDuration (the sum of the three duration variables) and AvgDurPerPage (TotalDuration divided by TotalPages, with zero page counts replaced by one to avoid division by zero). The final matrix had 29 columns. Sessions were split into a training set of 9,864 rows and a test set of 2,466 rows, using stratification on the target and a fixed random seed of 42. Numeric features were standardised with statistics learned from the training set only, so that no information from the test set leaked into scaling or clustering.

### 3.4 Unsupervised segmentation

K-Means (Lloyd, 1982) was fitted on the standardised training data with four clusters and ten random restarts. Each session was assigned to a cluster, and the assignment was one-hot encoded and appended to the feature matrix. Test sessions were assigned to the nearest training centroid. The number of clusters was kept small so that each segment could be profiled and explained to a business audience.

### 3.5 Stacked hybrid ensemble

Three base learners were chosen because they err in different ways. A Random Forest of 300 trees with balanced class weights provides variance reduction (Breiman, 2001). An XGBoost model with 300 trees, depth four, a learning rate of 0.05, row and column subsampling of 0.8 and a positive-class weight equal to the class ratio provides regularised gradient boosting (Chen &Guestrin, 2016). A LightGBM model with 300 trees, 31 leaves, a learning rate of 0.05 and balanced weights provides fast leaf-wise boosting (Ke et al., 2017). Their out-of-fold probabilities, generated by five-fold stratified cross-validation on the training set, were fed to a logistic-regression meta-learner, following the stacking principle of Wolpert (1992). All models were implemented in Python with scikit-learn (Pedregosa et al., 2011), XGBoost and LightGBM.

### 3.6 Baselines, threshold calibration and metrics

Six single classifiers served as baselines: logistic regression, a depth-limited decision tree, an RBF support vector machine, Random Forest, XGBoost and LightGBM. Two further variants isolate the hybrid design: the stacked ensemble without the clustering features, and the full hybrid model with a calibrated threshold. Because the positive class is the minority, the default threshold of 0.5 is rarely optimal. The calibrated threshold was selected only on the training set, by maximising F1 on out-of-fold predictions, and its value was 0.37. Models were scored on the untouched test set with accuracy, precision, recall, F1-score and the area under the ROC curve (AUC). Because accuracy is dominated by the majority class, F1 and AUC are treated as the primary indicators.

## 4. Results

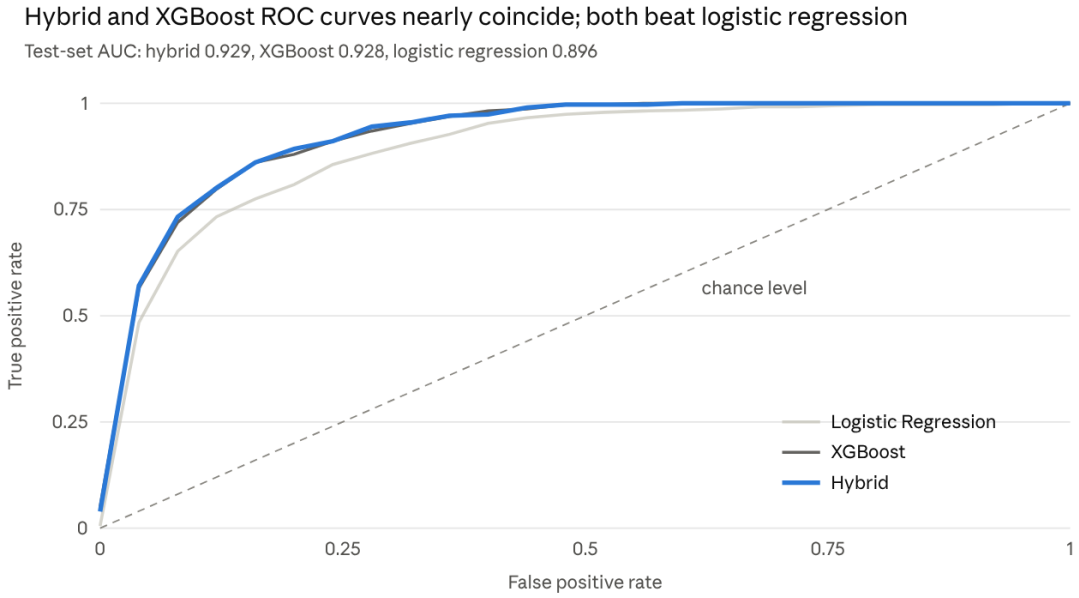
### 4.1 Model comparison

Table 3 reports test-set performance. Among the single learners, XGBoost achieved the best AUC (0.928), and Random Forest the best accuracy (0.898) but with a recall of only 0.524, so it missed nearly half of the buyers. The RBF support vector machine had the lowest recall (0.429). The hybrid model reached the highest AUC (0.929) and an accuracy of 0.899. The stacked ensemble without clustering performed almost identically (AUC 0.928, F1 0.661), which shows that the benefit comes mainly from combining the three tree learners. Applying the calibrated threshold to the hybrid model raised recall from 0.628 to 0.715 and F1 from 0.658 to 0.678, the best F1 of all configurations, at a modest cost in precision.

**Table 3. Test-set performance of all models (n = 2,466)**

| Model | Accuracy | Precision | Recall | F1 | AUC |
|---|---|---|---|---|---|
| Logistic Regression | 0.850 | 0.512 | 0.749 | 0.608 | 0.896 |
| Decision Tree | 0.852 | 0.514 | 0.812 | 0.629 | 0.895 |
| SVM (RBF) | 0.883 | 0.698 | 0.429 | 0.532 | 0.884 |
| Random Forest | 0.898 | 0.741 | 0.524 | 0.614 | 0.919 |
| XGBoost | 0.869 | 0.553 | 0.796 | 0.652 | 0.928 |
| LightGBM | 0.881 | 0.593 | 0.741 | 0.659 | 0.921 |
| Stacked ensemble (no clustering) | 0.900 | 0.698 | 0.628 | 0.661 | 0.928 |
| Hybrid (K-Means + stacked RF/XGB/LGB) | 0.899 | 0.690 | 0.628 | 0.658 | 0.929 |
| **Hybrid, tuned threshold (t = 0.37)** | 0.895 | 0.645 | 0.715 | **0.678** | **0.929** |

Figure 1 compares the ROC curves of the main models. The hybrid curve lies at or above the others over most of the range, although the differences among the tree-based models are small.




Experiment on UCI Online Shoppers Purchasing Intention data · hold-out test set, n = 2,466

**Figure 1.** ROC curves of logistic regression, XGBoost and the hybrid model on the hold-out test set (curves computed from the experiment in Section 3).

### 4.2 Error analysis

With the calibrated threshold, the hybrid model classified 1,934 non-buyers and 273 buyers correctly, produced 150 false alarms and missed 109 buyers (Table 4). In marketing terms, a false alarm costs one unnecessary incentive, while a missed buyer is a lost opportunity, so the lower threshold, which trades 5 percentage points of precision for about 9 points of recall, is reasonable when incentives are cheap relative to the value of a sale.

**Table 4. Confusion matrix of the hybrid model at threshold 0.37**

| | Predicted: no purchase | Predicted: purchase |
|---|---|---|
| Actual: no purchase | 1,934 | 150 |
| Actual: purchase | 109 | 273 |

### 4.3 Customer segments

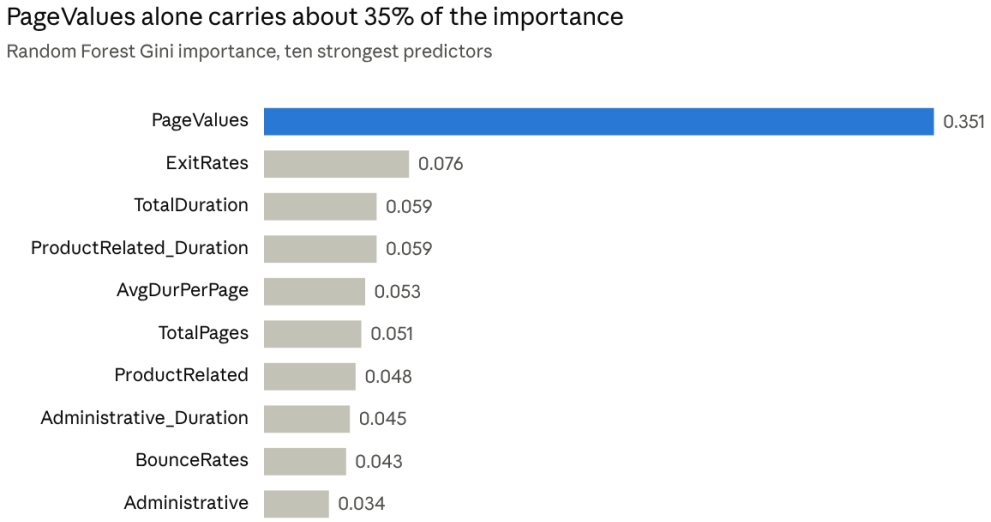
The K-Means step produced four segments, profiled in Table 5 by conversion in the training data. Segment 1 holds 1,040 training sessions with a conversion rate of 29.4%, nearly double the overall rate, while segment 0 holds 1,578 sessions that converted only 1.8% of the time. The large middle segment (7,172 sessions) converts at 16.4%, close to the average, and a very small segment of 74 sessions converts at 17.6%. The strong spread between the two extreme segments confirms that the clusters capture meaningful behavioural structure, even though they add little to predictive accuracy once the ensemble is present.

**Table 5. Conversion rate by K-Means segment (training set)**

| Segment | Sessions | Purchase rate |
|---|---|---|
| 0 | 1,578 | 1.8% |
| 1 | 1,040 | 29.4% |
| 2 | 7,172 | 16.4% |
| 3 | 74 | 17.6% |

### 4.4 Feature importance

A separate Random Forest was fitted to rank predictors. PageValues dominated with a Gini importance of 0.351, far above the next variables: ExitRates (0.076), TotalDuration (0.059), ProductRelated_Duration (0.059), AvgDurPerPage (0.053), TotalPages (0.051), ProductRelated (0.048), Administrative_Duration (0.045), BounceRates (0.043) and Administrative (0.034). Two of the three engineered features appear in the top six, which supports the decision to build aggregate engagement measures. Figure 2 displays the ranking.




Experiment on UCI Online Shoppers Purchasing Intention data · training set, n = 9,864

**Figure 2.** Ten most important predictors of purchase (Random Forest Gini importance, training set).

## 5. Discussion

### 5.1 What the results mean

The results lead to three observations. First, the visitor's trajectory through valuable pages matters far more than demographics or technical context. PageValues, an analytics measure of how much revenue a page has historically contributed, carries most of the signal, which agrees with the idea that purchase-ready visitors move toward checkout-oriented pages. Second, the hybrid design delivers the best ranking quality, but the margin over strong single boosters is slim; practitioners should not assume that added architectural complexity always pays. In this study the stacked ensemble was the real contributor and the clustering layer served chiefly as an interpretation device. Third, decision-threshold calibration produced a larger practical gain in F1 and recall than any change of algorithm, and it costs nothing to implement.

### 5.2 Managerial implications

A retailer could deploy the model as a scoring service that updates purchase probability while the session is in progress. Visitors in the high-conversion segment with a probability above the threshold could be left alone or shown a frictionless checkout, whereas mid-probability visitors, the most persuadable group, could receive a time-limited offer or social-proof messages such as recent-purchase notices. This is where the behavioural findings of Gaur et al. (2026) become useful: fear of missing out and the need for

social validation are precisely the levers that scarcity cues and review displays pull, and a prediction layer indicates whom to apply them to and whom to spare.

### 5.3 Limitations

Several limits should be noted. The evaluation uses one stratified split and no confidence intervals, so small differences, such as those between the hybrid and the plain stacked ensemble, should not be treated as statistically conclusive. The dataset comes from a single unnamed retailer and one year of traffic, so performance may not transfer to other sites or markets. The data record behaviour only; they contain no measures of fear of missing out, urgency or social influence, so the psychological mechanisms discussed in the literature are inferred rather than tested. The cluster count was fixed at four for interpretability rather than tuned. Finally, the system predicts association, not causation, and uplift modelling would be required to prove that an intervention changes outcomes.

## 6. Conclusion and Future Work

This paper developed and evaluated an intelligent consumer behaviour prediction system that joins K-Means segmentation with a stacked ensemble of Random Forest, XGBoost and LightGBM. On 12,330 real browsing sessions the hybrid achieved the best AUC (0.929) and, with a training-derived threshold, the best F1 (0.678) and a recall of 0.715. An ablation showed that stacking, not clustering, explains most of the predictive gain, while clustering offered clear, business-readable segments. Future work will add psychographic variables, such as survey-based measures of fear of missing out, instant gratification and social validation, to test whether they improve prediction beyond clickstream behaviour, will use repeated cross-validation with significance tests, and will examine sequence models and explainability tools such as SHAP values for individual-level explanations.

## Data Availability

The Online Shoppers Purchasing Intention dataset is publicly available from the UCI Machine Learning Repository (https://archive.ics.uci.edu/dataset/468) and from IEEE DataPort (DOI: 10.21227/e73k-cd23).

## References


1. Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32.
2. Chen, T., &Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). ACM.
3. Dietterich, T. G. (2000). Ensemble methods in machine learning. In *Multiple Classifier Systems* (pp. 1–15). Springer.
4. Gaur, H., Sehgal, S., Sharma, S., Gaur, P., &Kargeti, H. (2026). The Impact of FOMO, Instant Gratification, and Social Validation on Consumer Behavior in Online Marketing. *Cadernos De Pós-Graduação Em DireitoPolítico E Econômico*, *26*(1), 354–374. Retrieved from https://ceapress.org/index.php/cpgdpe/article/view/87
5. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, *30*, 3146–3154.
6. Lloyd, S. P. (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory*, *28*(2), 129–137.
7. Pedregosa, F., Varoquaux, G., Gramfort, A., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.
8. Sakar, C. O., Polat, S. O., Katircioglu, M., &Kastro, Y. (2019). Real-time prediction of online shoppers' purchasing intention using multilayer perceptron and LSTM recurrent neural networks. *Neural Computing and Applications*, *31*(10), 6893–6908.
9. Wolpert, D. H. (1992). Stacked generalization. *Neural Networks*, *5*(2), 241–259.