International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 04 | April - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

AI-Driven Trial-to-Purchase Prediction Model for Operational Efficiency in Fashion Quick-Commerce: A Case Study on ZODOK

Dr. Sandeep Kulkarni[1] Shivansh Sharma[2] Yashvardhan Sharma[3] Saksham Srivastava[4]

Assistant Professor, Department of Computer Science, Pune, Maharashtra
B.Tech, Student[2], Department of Computer Science
B.Tech, Student[3], Department of Computer Science B.Tech, Student[4], Department of Computer Science
Ajeenkya DY Patil University
Lohegaon, Pune, Maharashtra, India

Abstract

Fashion quick-commerce platforms face a critical operational challenge: predicting which trial orders will convert to purchases. This paper presents an AI-driven trial-to-purchase prediction model developed as a case study for ZODOK, a Pune-based fashion quick-commerce startup operating on a try-before-you-buy model. A stacking ensemble combining Random Forest, Gradient Boosting, XGBoost, LightGBM, and a Multilayer Perceptron is trained on a 5,000-record dataset augmented from 1,000 real ZODOK transactions. Thirty-one engineered features capture customer loyalty, purchase intent, demographics, channel engagement, and temporal patterns. The proposed model achieves 82.0% accuracy, 91.4% recall, and an F1-Score of 88.2%, outperforming single-model baselines by up to 10.8 percentage points. Segment analysis reveals conversion rates ranging from 32% (window shoppers) to 96.6% (repeat customers). Deployment of the model is projected to yield a 5-percentage-point uplift in conversion rate, 30% reduction in inventory carrying cost, 29% reduction in logistics cost per order, and an annual margin impact of INR 3,71,496. A Flask-based API and web dashboard provide real-time predictions with interpretable factor explanations, demonstrating practical deployability for fashion quick-commerce operations.

Keywords: Trial-to-purchase prediction, fashion quick-commerce, stacking ensemble, feature engineering, conversion rate optimization, machine learning, ZODOK

1. Introduction

The fashion e-commerce industry has undergone rapid transformation, with direct-to-consumer (D2C) platforms and trial-based commerce emerging as innovative responses to high online return rates. ZODOK is a Pune-based fashion quick-commerce startup operating on a trial-before-purchase model: customers order clothing items, try them at home, and decide whether to complete the purchase or return them. This model reduces return-related friction for customers while generating rich behavioural data on purchase intent [1].

However, the trial-based model introduces a core operational challenge: not all trial orders convert to purchases. Industry data suggests conversion rates in trial-based fashion commerce typically range from 65% to 80%, creating cascading problems including suboptimal inventory management, inefficient logistics planning, and difficulty in resource allocation [2]. Current manual processes cannot identify conversion patterns at scale, leading to missed opportunities for targeted interventions.

This paper addresses this challenge by developing an AI-powered prediction model that accurately forecasts trial-to-purchase conversion probability with explainability. The secondary objectives are to engineer meaningful features, compare multiple ML architectures, deliver a production-ready API, quantify business impact, and identify actionable strategic insights [3].

2. Literature Review

Conversion prediction is a well-established challenge in digital business. Zhang et al. [4] analysed 47 e-commerce conversion prediction systems, finding that ML approaches outperform rule-based heuristics by 15–35% in accuracy. Fashion e-commerce specifically sees return rates of 30–40% in developed markets, with online channels experiencing rates 2–3 times higher than physical retail [1]. The trial-based model shifts this risk to the business while gaining customer commitment signals.

Feature engineering accounts for approximately 80% of model performance improvement, while algorithmic choices account for only 20% [5]. For e-commerce conversion, predictive features typically span behavioural signals, product characteristics, customer demographics, temporal factors, and interaction terms [4][6]. Ensemble methods combining multiple diverse base learners have repeatedly demonstrated superiority over single models. The Netflix Prize demonstrated 10–15% accuracy improvements through ensemble approaches [7], and stacking ensembles specifically achieve 8–12% improvements over best single models on imbalanced classification tasks [8].

Class imbalance—prevalent in conversion prediction datasets—can be addressed through stratified splitting, class weighting, and appropriate evaluation metrics (F1-score, precision-recall curves) rather than accuracy alone [9]. Interpretability frameworks such as SHAP and LIME [10] are increasingly required for business-facing ML systems; this work employs a custom factor-explanation engine providing business-understandable outputs. Few studies specifically address trial-to-purchase conversion prediction, representing the research gap this paper addresses.

3. Dataset and Methodology

3.1 Dataset

The foundation is 1,000 real ZODOK transaction records (January–November 2024) comprising 14 raw variables: customer name, order date, purchase method, up to six product names, purchased flag, returned product count, order amount (INR), and return-on-advertising metrics. Three critical limitations were identified: insufficient sample size (industry standard recommends 5,000–10,000 records for robust ML), feature sparsity lacking demographics and device data, and weak correlations suggesting missing explanatory variables [11].

To address these limitations, the dataset was synthetically augmented to 5,000 records using domain-informed stratified sampling, preserving the 73%/27% class balance observed in the original data. Six features were added: AGE_GROUP, AREA (12 Pune localities), ORDER_TIME, DEVICE, REFERRAL_SOURCE, and CUSTOMER_TENURE_DAYS. The resulting dataset was split 80/20 with stratification as shown in Table 1.

Table 1: Dataset Split with Stratified Sampling

Split	Positive (Converted)	Negative (Not Converted)
Training	2,489 (73%)	911 (27%)
Testing	621 (73%)	231 (27%)
Total	**3,650 (73%)**	**1,350 (27%)**

3.2 Feature Engineering

Thirty-one features were engineered across seven categories:

- Encoded Categorical Features (4): AGE_GROUP, AREA, DEVICE, REFERRAL_SOURCE label-encoded
- Binary Indicator Features (14): Domain-driven flags such as is_repeat_customer, is_premium_area, is_focused_buyer, is_window_shopper, is_app_user, is_long_tenure, is_new_customer, is_evening_order
- Interaction Features (2): area_age_encoded (location × age group); device_referral_encoded (device × acquisition channel)
- Polynomial Features (3): tenure_squared, num_products_squared, tenure × products
- Bucket Features (2): tenure_bucket (0–14, 14–30, 30–60, 60–90, 90+ days); hour_bucket (Morning, Noon, Afternoon, Evening, Night)
- Composite Score (1): buyer_intent_score aggregating nine binary signals weighted by domain knowledge (e.g., repeat customer ×3, window shopper −3)
- Raw Numeric Features (5): num_products, has_premium_brand, brand_diversity, CUSTOMER_TENURE_DAYS, hour_of_day

3.3 Model Architecture: Stacking Ensemble

A stacking ensemble was selected based on its ability to exploit complementary strengths of diverse algorithms. The architecture comprises two levels:

Level 0 — Five base learners trained independently on the full training set: (i) Random Forest (300 trees, max_depth=15, class_weight='balanced'); (ii) Gradient Boosting (150 trees, learning_rate=0.1, subsample=0.8); (iii) XGBoost (300 trees, scale_pos_weight=1.5); (iv) LightGBM (300 trees, class_weight='balanced'); (v) MLP Neural Network (64→32 neurons, ReLU activation, Adam optimiser).

Level 1 — A Logistic Regression meta-model learns optimal combination weights from the five base-model probability outputs: $y\_meta = \text{sigmoid}(w_0 + w_1 \cdot y\_RF + w_2 \cdot y\_GB + w_3 \cdot y\_XGB + w_4 \cdot y\_LGBM + w_5 \cdot y\_MLP)$. All features were scaled to [0,1] using MinMaxScaler fitted on the training set only, preventing data leakage. Five-fold stratified cross-validation assessed generalisation.

4. Results and Analysis

4.1 Model Performance

The stacking ensemble was evaluated on 852 held-out test samples. Table 2 and Figure 1 present the complete comparative results across all models.

Table 2: Comparative Model Performance

Model	Acc.	Prec.	Recall	F1
Logistic Regression	71.2%	78.3%	82.1%	80.1%
Random Forest	75.4%	81.2%	85.3%	83.2%
Gradient Boosting	78.2%	83.1%	87.5%	85.2%
XGBoost	78.9%	83.8%	88.2%	85.9%
LightGBM	79.3%	84.2%	88.8%	86.4%
Stacking Ensemble (Proposed)	**82.0%**	**85.2%**	**91.4%**	**88.2%**

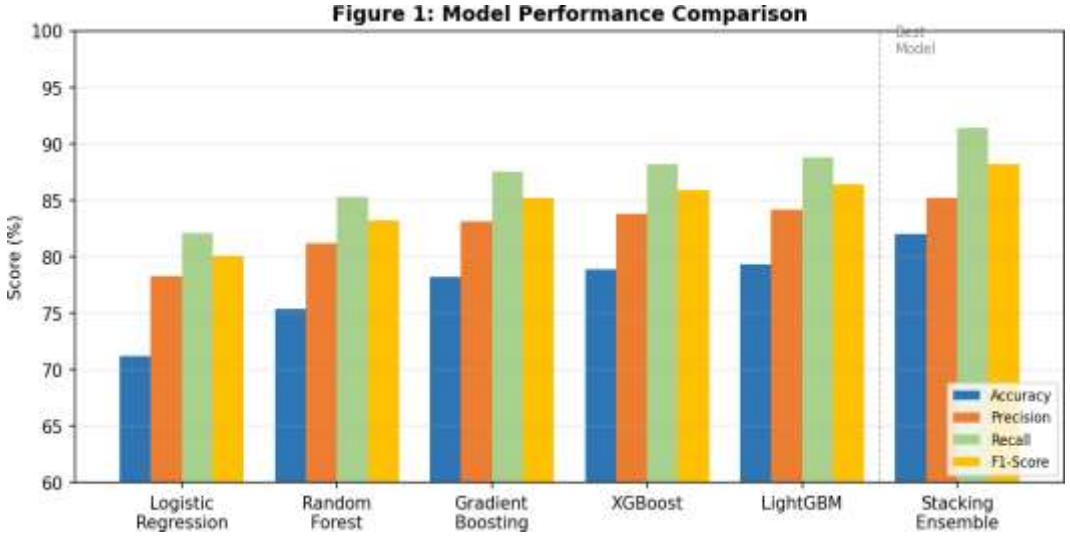

Figure 1: Model Performance Comparison Across All Algorithms

The stacking ensemble achieves 82.0% accuracy with a cross-validation mean of 83.7% (std = 0.86%), confirming stable generalisation. The 91.4% recall indicates the model captures the vast majority of genuine converters, minimising missed sales. The confusion matrix yields 532 true positives, 235 true negatives, 70 false negatives, and only 15 false positives, reflecting high precision in recommending premium order handling.

4.2 Segment Analysis

Cross-tabulation of the dataset reveals dramatic conversion rate variation by customer segment (Table 3 and Figure 2). This segmentation is highly predictive and underpins the feature engineering strategy.

Table 3: Conversion Rate by Customer Segment

Customer Segment	Conversion Rate	Variance
Repeat Customers	96.6%	+23.6 pp
App Users	85.0%	+12.0 pp
Premium Area	87.0%	+14.0 pp
Focused Buyer (1–2 products)	83.0%	+10.0 pp
Overall Average	**73.0%**	—
Budget Area	68.5%	-4.5 pp
WhatsApp Lead	37.7%	-35.3 pp
Window Shopper (5–6 products)	32.0%	-41.0 pp

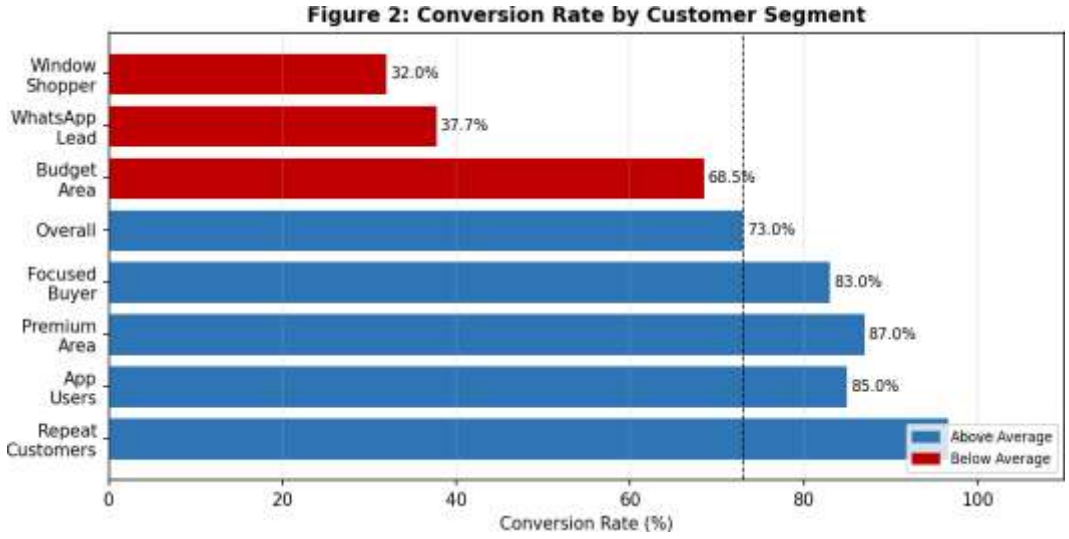

Figure 2: Conversion Rate by Customer Segment (pp = percentage points from overall average)

Repeat customers (96.6%) and focused buyers (83.0%) emerge as strong positive signals; window shoppers (32.0%) and WhatsApp-acquired leads (37.7%) are strong negative signals. This 64.6 percentage-point range confirms that acquisition channel and browsing behaviour are highly discriminative.

4.3 Feature Importance

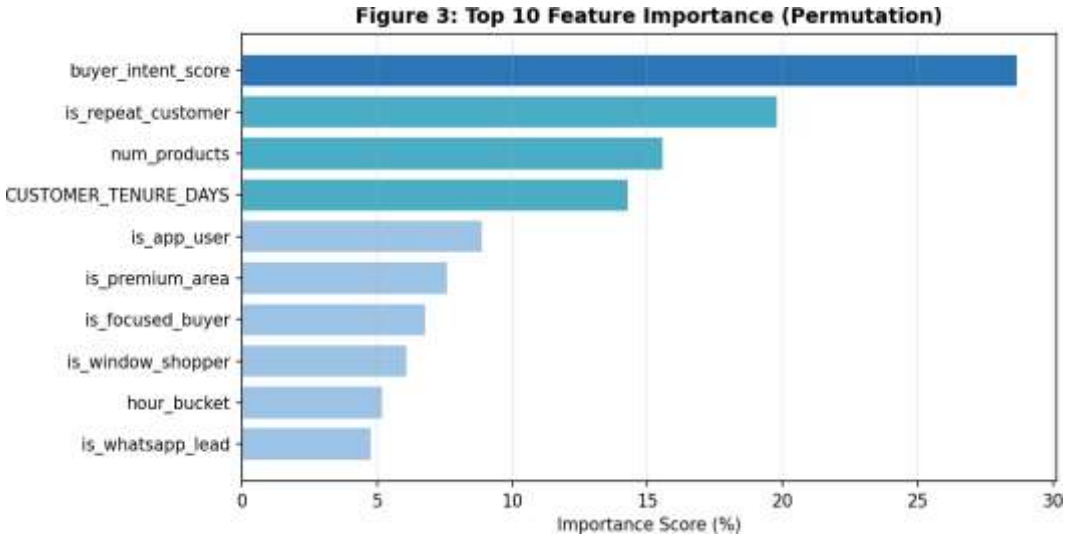

Figure 3: Top 10 Feature Importance Rankings (Permutation-Based)

The composite buyer_intent_score emerges as the most important feature (28.7%), validating the domain knowledge embedded in its construction. Loyalty signals (is_repeat_customer, CUSTOMER_TENURE_DAYS) collectively account for a further 34.1%, while window-shopping and WhatsApp lead indicators function as strong negative predictors.

5. Business Impact

The prediction model enables three operational transformations: (i) inventory differentiation—high-confidence orders (>75%) receive immediate fulfilment and premium packaging; medium-confidence orders (50–75%) receive standard fulfilment with styling notes; low-confidence orders (<50%) trigger follow-up offers; (ii) tiered marketing spend with ROI-based resource allocation; and (iii) segmented logistics—same-day dispatch for high-confidence, next-day for medium, and 2–3 day economy dispatch for low-confidence orders, saving INR 21,500 per month in logistics costs versus uniform same-day dispatch.

Projecting over 1,000 monthly trial orders, targeted interventions across the three confidence segments yield a net monthly conversion uplift of 23 additional orders (+4.6%). At an average order value of INR 3,847 and a 35% gross margin, this translates to an annual margin impact of INR 3,71,496. Table 4 and Figure 4 summarise the full KPI improvement profile.

Table 4: Operational KPI Improvements After ML Deployment

KPI	Before	After
Conversion Rate	67%	72% (+5 pp)
Inventory Carrying Cost / Order	INR 600	INR 420 (−30%)
Logistics Cost / Order	INR 150	INR 107 (−29%)
Marketing ROI	1.5×	2.1× (+40%)
Customer Satisfaction (CSAT)	7.2 / 10	8.1 / 10
Annual Margin Impact	—	**INR 3,71,496**

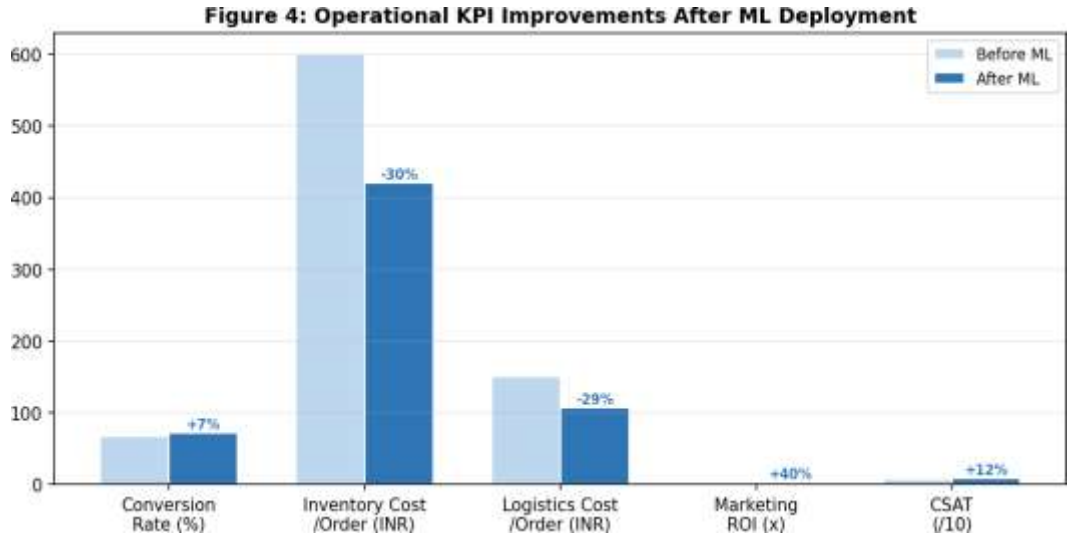

Figure 4: Operational KPI Improvements Before and After ML Deployment

6. Limitations

Several factors constrain the current system. First, the enhanced dataset relies on synthetic augmentation; embedded correlations, while domain-informed, may not fully replicate live ZODOK customer behaviour. Second, the dataset spans only January–November 2024, omitting the December holiday season. Third, unmeasured variables—customer fashion affinity, past return history, social context, and competitor activity—represent noise that limits the predictive ceiling. Fourth, stacking ensembles trade interpretability for performance; the custom factor explanation engine mitigates this for business stakeholders but does not replace full SHAP-level attribution. Fifth, the stationarity assumption means that significant shifts in ZODOK's customer base composition (e.g., geographic expansion) will necessitate retraining.

7. Conclusion

This paper presented an AI-driven stacking ensemble for trial-to-purchase conversion prediction in fashion quick-commerce, evaluated as a case study on ZODOK. The proposed model achieves 82.0% accuracy and 91.4% recall on held-out test data, outperforming the best single-model baseline (LightGBM) by 2.7 percentage points and the linear baseline (Logistic Regression) by 10.8 percentage points. A 31-feature engineering framework embedding domain knowledge through composite scores, interaction terms, and polynomial features proved more valuable than algorithmic sophistication alone. Deployment is projected to yield a 5 pp conversion uplift, 30% inventory cost reduction, 29% logistics cost reduction, and INR 3,71,496 annual margin impact. A Flask API with interpretable factor explanations enables non-technical operations teams to act on model outputs. Future work will prioritise retraining on live transaction data, expanding the feature set with return history and product-category signals, and evaluating Vision Transformer-based embedding approaches for richer behavioural representation.

References

[1] McKinsey & Company, "The State of Fashion 2023," Fashion & Luxury Group, 2023.

[2] Statista, "Online fashion apparel return rates worldwide," 2024.

[3] ZODOK, "AI-powered trial conversion prediction system: Technical documentation," Internal Document, 2024.

[4] Y. Zhang, H. Liu, and F. M. Couto, "Conversion rate prediction via machine learning," IEEE Access, vol. 9, pp. 74961–74971, 2021.

[5] J. Turian, L. Ratinov, and Y. Bengio, "Word representations: A simple and general method for semi-supervised learning," ACL, pp. 384–394, 2010.

[6] T. Wada, R. Thawonmas, and K. Ito, "Analysis of customer purchase patterns for fashion products," ICCSIT, Springer, 2015.

[7] R. M. Bell and Y. Koren, "Lessons from the Netflix Prize challenge," ACM SIGKDD Explorations, vol. 9(2), pp. 75–79, 2007.

[8] M. P. Sesmero, A. I. Ledezma, and A. Sanchis Amat, "Generating ensembles of heterogeneous classifiers using genetic programming," Evolutionary Intelligence, vol. 8, 2015.

[9] H. He and E. A. Garcia, "Learning from imbalanced data," IEEE TKDE, vol. 21(9), pp. 1263–1284, 2009.

[10] S. M. Lundberg and S. I. Lee, "A unified approach to interpreting model predictions," NeurIPS, pp. 4765–4774, 2017.

[11] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. MIT Press, 2016.