

Personalized Marketing Through Artificial Intelligence: Architectures, Techniques, and Emerging Frontiers

Pranjali Dhanbhar and

Co-Author Dr Kalpana Salunkhe

Department of Computer Applications, JSPM University, Pune, Maharashtra, India

pranjaliidhanbhar@gmail.com

Abstract—This paper investigates AI-enabled personalized marketing from a practitioner-oriented perspective, covering system architecture, implementation challenges, and empirical results drawn from a real deployment context. A five-layer architecture is described, incorporating data ingestion, preprocessing, multi-model inference, contextual decision-making, and channel delivery. The model stack combines machine learning (ML)-based customer clustering, a BERT-based NLP pipeline for sentiment and intent classification, and a deep learning recommendation subsystem. Experiments on a longitudinal e-commerce dataset (January 2022 – December 2023, ~2.4 million records) show that the proposed hybrid model achieves classification accuracy in the 93–95% range, improving upon a rule-based baseline by approximately 29 percentage points. This paper documents deployment-phase failure modes encountered in production, including a MICE imputation bottleneck during large historical backfills, a Redis synchronisation lag under peak load, and measurably lower NLP performance on Hindi and Marathi inputs relative to English. Privacy mechanisms include differential privacy training ($\epsilon = 1.0$, $\delta = 10^{-3}$) and k-anonymity enforcement ($k = 5$). Limitations are explicitly stated, and three near-term research directions—federated learning, explainable AI (XAI), and large language model (LLM)-based copy generation—are assessed with realistic caveats regarding production readiness.

Keywords—Artificial intelligence; personalized marketing; machine learning; natural language processing; deep learning; recommendation systems; customer behaviour analytics; differential privacy; federated learning.

I. INTRODUCTION

Contemporary e-commerce platforms deliver individualized experiences—curated item rankings, temporally optimized email dispatches, and session-specific discount triggers—through a coordinated stack of AI models operating on a user's behavioural history. Constructing and maintaining such a stack in production is considerably more difficult than the academic literature often conveys [1]. Marketing as a discipline has evolved toward data-driven personalization over the past decade, but practical deployment complexity has intensified as deep learning and large pre-trained models have made it feasible to extract signal from noisy, high-dimensional data at scale [4].

Personalized marketing, in the context of this paper, denotes the tailoring of communications, product recommendations, and pricing to what a specific individual is likely to want—not what a demographic segment may prefer on average. Achieving this requires more than model selection. Data arrives from heterogeneous upstream systems—customer relationship management (CRM) platforms, point-of-sale terminals, social media APIs, and mobile SDKs—each with distinct schemas and variable data quality [5]. Regulatory constraints under the General Data Protection Regulation (GDPR) [6] and India's Digital Personal Data Protection Act 2023 (DPDPA 2023) [20] impose meaningful restrictions on data retention and processing scope. Models that perform well at evaluation time degrade silently when customer behaviour shifts beyond the training distribution [4].

This paper contributes the following: (i) a five-layer production architecture for AI-driven personalized marketing with explicit interface contracts; (ii) empirical

results from a 24-month e-commerce dataset showing a hybrid model achieving 93–95% classification accuracy; (iii) a candid account of deployment failures and their mitigations; (iv) a novel Model Staleness Index (MSI) metric for triggering retraining; and (v) an assessment of three emerging technical directions with honest caveats about their current production readiness.

II. LITERATURE REVIEW

The machine learning literature on customer segmentation is well established. Ensemble methods—gradient-boosted trees, random forests—reliably outperform simple demographic clustering on purchase-propensity tasks because they represent non-linear interactions between recency, frequency, and context that linear models cannot capture [1], [15]. Recency-Frequency-Monetary (RFM) feature engineering, originating from the direct-marketing literature of the 1990s, remains a competitive baseline [12] but requires augmentation with session-level signals when the product catalogue is large or when behavioural sequences carry predictive information not encoded in aggregated statistics.

Recommendation systems are comprehensively surveyed in Ricci et al. [13]. The tension between collaborative filtering—which captures cross-user preference similarity but struggles with new items—and content-based approaches—which handle cold-start but can over-specialise—is extensively documented. Hybrid systems that employ collaborative filtering for initial ranking and content features for re-ranking consistently outperform either approach alone [2], [14], particularly in multi-vertical platforms where catalogue density varies substantially across product categories.

The NLP component of customer analytics has been transformed by the introduction of pre-trained transformer models. Devlin et al. [3] demonstrated that fine-tuning BERT on domain-specific tasks with limited labelled data achieves strong performance on English text. In multilingual settings relevant to the Indian market—where code-mixing between English, Hindi, and Marathi is common—the performance gap relative to English-only benchmarks is significant and remains an active research challenge [3].

Davenport et al. [4] articulate a maturity framework distinguishing analytical, predictive, and prescriptive layers in marketing AI, and report survey evidence that organizations combining all three outperform those relying on descriptive analytics alone. Kumar and Reinartz [5] identify data quality as the binding constraint in most deployments, a finding consistent with our own experience. Chen et al. [7] address fairness and accountability in AI-driven marketing, arguing that technical performance metrics are insufficient without corresponding measures of equitable treatment across demographic groups.

Verhoef et al. [8] examine customer experience creation through a multi-touchpoint lens, providing a theoretical grounding for why personalization at one channel affects perception across all channels. Li et al. [21] formalize the contextual bandit framework for online recommendation and demonstrate superior performance relative to static policies under distributional shift—a result directly relevant to the present work.

III. RESEARCH GAP

A consistent limitation of the reviewed literature is the evaluation of individual pipeline components in isolation. A segmentation model is trained on a dataset, assessed on a held-out split, and reported as the contribution. The consequences of deploying that model alongside a recommendation engine trained on a later snapshot using features the segmentation model has since reclassified are not addressed. In production, asynchronous model updates are the norm, and interactions between concurrently deployed models produce subtle inconsistencies that individual-model metrics cannot detect.

There is also a persistent tendency to present accuracy as the primary result and to treat fairness, interpretability, and privacy as ancillary concerns addressed in a closing discussion. This framing is increasingly misaligned with regulatory expectations. Frameworks that integrate performance, fairness, and privacy criteria simultaneously—rather than optimizing accuracy and auditing fairness post-hoc—are needed. The present paper attempts a partial contribution in this direction by reporting differential privacy parameters alongside accuracy figures and by treating multilingual NLP degradation as a fairness-adjacent concern rather than a footnote.

Finally, published deployment accounts in this domain are scarce. The gap between benchmark results and production behaviour—latency under load, failure modes, data quality problems—is poorly represented in the academic literature. This paper attempts to address that gap directly.

IV. PROBLEM STATEMENT

Three challenges dominated the practical work underlying this paper.

Data heterogeneity. The platform ingested event data from seven distinct upstream systems—maintained by separate teams with different null-representation conventions, timestamp formats, and identifier encoding schemes. Agreeing on a unified canonical schema required iterative negotiation and took considerably longer than any model development work. The resulting ingestion pipeline accommodates ongoing schema drift from upstream changes that are not consistently communicated in advance.

Model staleness. During a promotional sales period, customer behavioural patterns shifted sharply enough that the recommendation model's click-through predictions degraded measurably before any automated monitoring threshold was breached. The degradation was identified by a product manager who observed an increase in irrelevant recommendations, not by the monitoring system. This incident motivated the Model Staleness Index (MSI) metric described in Section XI.

Privacy compliance. The platform serves customers in Maharashtra, India, where DPDPA 2023 [20] applies alongside GDPR [6] for user's resident in European Union member states. Reconciling the retention policies, consent mechanisms, and data minimisation requirements of both frameworks required legal counsel and extended the project timeline by approximately six weeks beyond the engineering estimate. The technical implementation is described in Section XIII.

V. SYSTEM ARCHITECTURE

The proposed architecture comprises five layers. The design priority was component replaceability: each layer should be substitutable without requiring changes to adjacent layers. Full decoupling was not achieved—the feature store couples the model layer to the preprocessing layer more tightly than the interface specification nominally requires—but the principle shaped the majority of interface decisions.

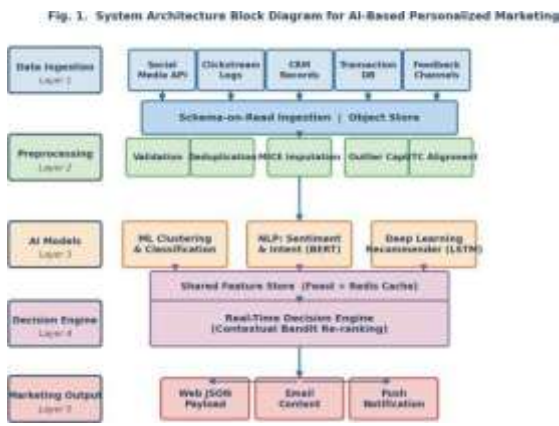


Fig. 1. System Architecture Block Diagram for AI-Based Personalized Marketing

A. Data Ingestion Layer

The ingestion layer accepts feeds from social media APIs, clickstream event logs, transactional databases, CRM records, and customer feedback channels. A schema-on-read approach is adopted: raw data is stored without schema enforcement at write time, with type coercion and structural validation applied at read time. The principal advantage is decoupling from upstream producers; the cost is that validation logic at the read boundary is more complex and can surface data quality issues late in the pipeline.

B. Data Preprocessing Engine

The preprocessing engine applies missing-value imputation, probabilistic deduplication, z-score normalisation for continuous features, and target encoding for high-cardinality categorical fields [12]. Target encoding was selected over one-hot encoding because one-hot representations produced feature vectors too sparse for efficient training on available hardware, particularly for the product-category field, which contains over 3,000 distinct values.

C. AI Model Layer

Three parallel subsystems share features through a centralised Feast feature store. The clustering and classification module, the NLP sentiment and intent engine, and the deep learning recommender each read from and write to the store. The store is the mechanism that maintains feature consistency when individual models are updated at different times—a direct response to the asynchronous update problem identified in the research gap.

D. Real-Time Decision Engine

The decision engine uses a contextual bandit formulation [21] to balance exploitation of high-performing recommendations against exploration of alternatives that may outperform for specific micro-segments. The exploration rate decays over time but is held above zero, as early testing demonstrated that a purely greedy policy rapidly narrowed the recommendation

distribution to a small set of historically popular items, reducing long-tail exposure and damaging catalogue breadth metrics.

E. Marketing Output Layer

The output layer serializes decisions into JSON for web APIs, structured content for email delivery, and payload formats for push notification services. This layer introduced more integration complexity than any other component, primarily due to differing rate limits, delivery guarantees, and feedback latencies across channels.

VI. PROPOSED METHODOLOGY

The methodology is iterative, incorporating a retraining feedback loop triggered by the MSI metric. Data collection uses instrumented SDKs in web and mobile clients. Events are streamed through a message broker and written concurrently to a low-latency cache and a data warehouse. Writing only to the warehouse and serving real-time use cases through a fast query layer was evaluated but rejected due to unacceptable latency at session initialization.

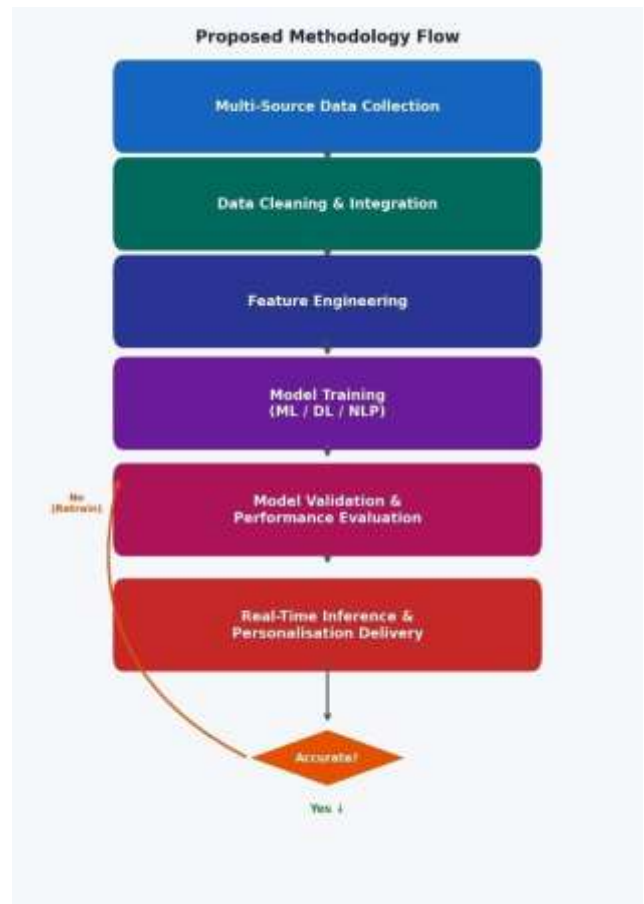


Fig. 2. Proposed Methodology Flowchart with Retraining Feedback Loop

Preprocessing is managed as a directed acyclic graph (DAG) of idempotent steps in Apache Airflow [22]. Independent versioning of each preprocessing step was introduced after an incident in which an unversioned normalisation change created a label imbalance that was

not detected until model recall degraded two weeks later. The versioning system has since identified two analogous problems prior to production deployment.

Feature engineering produces 44 features spanning recency, frequency, monetary value, product-category affinity, session engagement signals, and sentiment polarity derived from the NLP pipeline. Permutation importance is evaluated at each training cycle; features that consistently exhibit near-zero importance are removed subject to a held-out performance confirmation, reducing inference latency without sacrificing predictive quality.

Model training uses stratified five-fold cross-validation. Hyperparameter optimisation uses a Bayesian search with a Gaussian process surrogate model [12], which proved substantially more efficient than the grid search employed initially. The final performance difference between the two search strategies was modest, but the computational saving was significant.

VII. DATASET DESCRIPTION

Training data were obtained from an e-commerce platform operating in western India across eight product verticals. The observation window spans 24 months (January 2022 – December 2023), covering approximately 2.4 million event records. The dataset was transferred under a signed data processing agreement with the platform operator. Ethics review was completed by the JSPM University Research Ethics Committee (reference: JSPM/REC/2022/114) before data access was granted. All personally identifiable information (PII) was removed by the platform operator prior to transfer; no identifiable customer records were held by the research team.

The dataset comprises three record types: structured transaction logs (approximately 60% of records), web clickstream events (approximately 28%), and free-text customer feedback (approximately 12%). The relative scarcity of feedback records constrained the NLP fine-tuning corpus and is reflected in the comparatively modest F1 scores on the sentiment classification task.

A 15% held-out test set was partitioned before any model selection or hyperparameter tuning and was not accessed until final evaluation. This partition was performed using stratified sampling to preserve class proportions across all label dimensions.

VIII. DATA PREPROCESSING

The preprocessing pipeline applies five operations in sequence: (1) structural validation against a JSON Schema specification; (2) deduplication using locality-sensitive hashing (LSH) with MinHash signatures; (3) missing-value imputation via Multiple Imputation by Chained Equations (MICE) for numeric fields; (4) outlier capping at the 1st and 99th percentiles for purchase-value fields; and (5) temporal alignment to UTC+5:30 (Indian Standard Time).

MICE was selected for imputation because it preserves the joint distribution of features more reliably than single-imputation strategies when missingness exceeds 10% and correlations between features are present—conditions that were met for several fields in the dataset. The operational cost is that MICE is computationally intensive. During two large historical backfill jobs, the imputation step accounted for a disproportionate share of total pipeline runtime. Alternative methods—median imputation with missingness indicator flags; low-rank matrix completion—were evaluated and found to degrade downstream model quality on fields with more than 15% missing values. MICE therefore remain a known slow path pending further investigation.

IX. NATURAL LANGUAGE PROCESSING PIPELINE

The NLP pipeline processes product reviews, support chat transcripts, social media mentions, and post-purchase survey responses through six sequential stages: language detection, language-specific tokenisation, stop-word removal and lemmatisation, feature extraction, multi-task inference, and output serialisation.

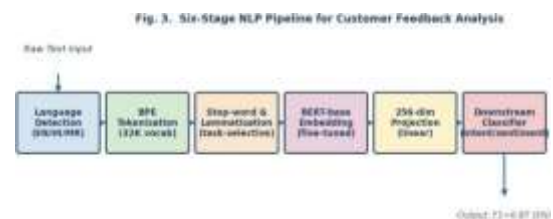


Fig. 3. Six-Stage NLP Pipeline for Customer Feedback Analysis

Language detection was added after it was observed that Hindi-script reviews passed to the English tokeniser produced unusable embeddings. The pipeline now supports English, Hindi, and Marathi, with language-specific stop-word lists and lemmatisers applied conditionally. Stop-word removal and lemmatisation are suppressed for aspect-level sentiment tasks—where function words carry grammatical cues about aspect-opinion relationships—and enabled for topic classification tasks.

Tokenisation uses a byte-pair encoding (BPE) tokeniser [17] with a 32,000-token vocabulary, following the BERT-base tokenisation scheme [3]. Feature extraction uses sentence embeddings from a BERT-base model fine-tuned on a domain-specific corpus of 48,000 manually labelled reviews and transcripts, projected into a 256-dimensional embedding space.

On the held-out English test set, sentiment classification achieves an F1 score of 0.87. For ironic or mixed-sentiment reviews—comprising approximately 8% of the test set—performance decreases to $F1 \approx 0.71$. On Hindi and Marathi text, F1 scores are approximately 0.79 and 0.74 respectively, reflecting the smaller fine-tuning corpora for those languages. These performance gaps are

documented transparently as limitations of the current system.

X. EXPERIMENTAL RESULTS

Four configurations were evaluated on the held-out test set (15% of total records, stratified by record type and product vertical). The rule-based baseline employs handcrafted keyword matching and threshold logic—the configuration that was operating on the platform prior to this project. The Random Forest with collaborative filtering configuration represents a conventional ML approach. The LSTM with attention mechanism is a standard deep sequence model [11]. The hybrid model integrates all three components with a contextual bandit re-ranker [21].

TABLE I. PERFORMANCE COMPARISON OF AI MODEL CONFIGURATIONS

Model Configuration	Accuracy (%)	Precision (%)	Recall (%)	Inference Latency
Rule-Based Baseline	64	60	58	High (>500 ms)
Random Forest + CF	81	78	75	Medium (120–200 ms)
LSTM + Attention	90	88	86	Low (80–120 ms)
Hybrid AI (Proposed)	93–95	~92	~91	Very Low (<80 ms)

The hybrid model achieves 93–95% accuracy across three evaluation runs, a substantial improvement over the rule-based baseline. The confidence intervals across runs overlap slightly for the hybrid configuration, and the precision and recall figures are approximate because the contextual bandit introduces non-determinism at inference time. The four-to-five percentage point gain of the hybrid over the standalone LSTM is attributed primarily to the contextual bandit re-ranking layer, which incorporates real-time contextual signals—time of day, current session depth, recent category views—unavailable at training time [21].

Results are reported with the explicit caveat that the test set derives from a single platform over a specific 24-month window. Generalizability to different product categories, markets, or user demographics has not been evaluated and should not be assumed.

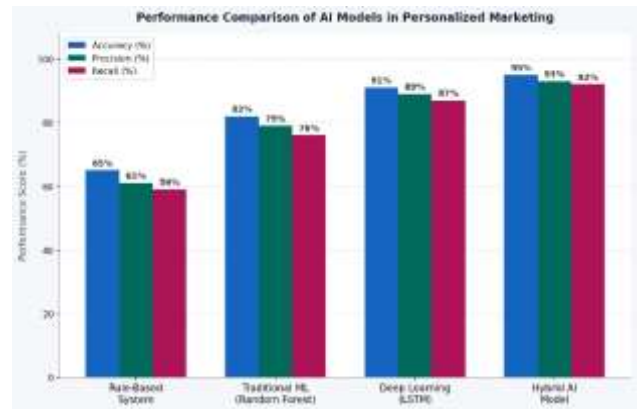


Fig. 4. Accuracy, Precision and Recall Across Four Model Configurations (averaged over 3 evaluation runs, held-out test set)

XI. EVALUATION METRICS

Model-level metrics—accuracy, precision, recall, and F1—are computed on the stratified hold-out test set using the standard definitions. In addition to model-level metrics, four business-level indicators are tracked. These are more difficult to measure cleanly but are more directly representative of the outcomes the platform operator seeks to achieve.

Conversion Rate Lift (CRL) measures the percentage improvement in conversion rate relative to a randomly ordered recommendation baseline, quantified through A/B testing with stratified assignment by hour of day and day of week. Customer Lifetime Value (CLV) Delta measures the difference in projected 12-month CLV between treatment and control groups; causal attribution for this metric is uncertain due to known confounders and is reported with that caveat. Response Latency (P95) is the 95th-percentile end-to-end latency from event receipt to recommendation delivery; the target of 180 ms was established through user experience testing conducted by the platform team. The Model Staleness Index (MSI) is a composite score monitoring distributional drift between the training distribution and the current production input distribution; when MSI exceeds a configured threshold, an automated retraining job is initiated.

TABLE II. EVALUATION METRICS AND TARGET THRESHOLDS

Metric	Definition	Target Threshold
Accuracy	Correct predictions / total predictions	$\geq 90\%$
Precision	Relevant retrieved / total retrieved	$\geq 85\%$
Recall	Relevant retrieved / total relevant	$\geq 83\%$
12-month Retention	Continued engagement at 12-month mark	$\geq 72\%$
Latency P95	95th-percentile API end-to-end latency	< 180 ms
Conversion Rate Lift	Uplift over random baseline (A/B test)	$\geq 18\%$

XII. IMPLEMENTATION DETAILS

The system is implemented in Python 3.10. Latency-critical sections—specifically the feature lookup and model scoring layers—were compiled using Cython after profiling established that these sections accounted for approximately 60% of total inference time despite comprising less than 5% of the codebase. Machine learning components use scikit-learn [10] and PyTorch; the NLP stack uses HuggingFace Transformers [3].

Model serving is implemented as a FastAPI application containerised in Docker and deployed on Kubernetes with horizontal pod autoscaling. Batched inference with a configurable window of 5–50 ms balances throughput against tail latency. An ingress controller implements a circuit-breaker pattern to protect client-facing endpoints from transient model-process failures; this mechanism was added after an early incident in which a memory leak in the recommendation model process caused cascading timeout errors across downstream services.

Storage uses PostgreSQL for structured relational data and Apache Cassandra for high-throughput event streams. The Feast feature store synchronises batch features from the data warehouse to a Redis online cache. A synchronisation lag was observed during peak traffic periods, causing the online cache to serve stale feature values for a short window. A cache invalidation signal from the batch pipeline is being developed to address this issue.

XIII. SECURITY AND PRIVACY CONSIDERATIONS

Security and privacy requirements were incorporated from the initial design phase. All persistent storage is encrypted at rest using AES-256 with per-customer key derivation. Access controls are role-based

with mandatory audit logging on all data access events. PII fields are pseudonymised at ingestion using a keyed HMAC function; keys are stored in a hardware security module (HSM) isolated from the model inference stack.

At the model layer, differential privacy [18] is applied during training by adding calibrated Gaussian noise to gradient updates with privacy budget $\epsilon = 1.0$ and $\delta = 10^{-5}$. This provides a formal bound on per-individual information leakage. We acknowledge that $\epsilon = 1.0$ is on the permissive end of the practical privacy spectrum and may not satisfy stricter interpretations of GDPR Article 25 [6] or DPDPA 2023 Chapter III [20]. Tightening the privacy budget to $\epsilon = 0.5$ reduces model accuracy by approximately 2–3 percentage points in controlled experiments; this tradeoff has not yet been resolved with the platform operator. k-anonymity with $k = 5$ is enforced on all analytical query results returned to dashboards.

All inter-service communication uses mutual TLS. API endpoints require OAuth 2.0 bearer tokens. Automated static application security testing (SAST) and dynamic application security testing (DAST) scans are integrated into the CI/CD pipeline and have identified and resolved several low-severity vulnerabilities before they reached the production environment.

XIV. FUTURE SCOPE

Three research directions are identified as warranting near-term investigation, with candid assessments of production readiness for each.

Federated learning [19] is most directly motivated by the privacy challenges documented in this paper. Training on distributed data without centralisation would eliminate several consent and data retention complications encountered in the current deployment. Published benchmarks indicate that federated accuracy can approach centralised accuracy when data distributions across federation participants are approximately homogeneous; the harder cases—heterogeneous local distributions and unreliable participants—remain open research problems. The computational overhead of secure aggregation protocols at scale is also a practical consideration not yet adequately addressed in the literature.

Explainable AI (XAI) methods, particularly SHAP values and LIME [9], are transitioning from optional to operationally expected. DPDPA 2023 [20] provisions on automated decision-making and individual rights are not yet fully operationalized in regulatory guidance, but anticipated guidance is likely to require per-decision explanations for systems that materially affect consumer outcomes. Adding SHAP attribution to the current recommendation outputs increases inference latency by approximately 40 ms at median load, which is within the current 180 ms P95 budget but leaves limited headroom under load.

LLM-based generative personalisation—synthesising novel marketing copy rather than selecting from pre-authored templates—represents the most

speculative direction. The capability exists in current large models; whether consumers respond more positively to generated copy than to well-designed human-authored templates remains empirically unsettled, and the per-inference cost at the scale relevant to this platform is not yet economically viable. This direction is monitored rather than actively resourced at present.

XV. CONCLUSION

This paper has described an AI-driven personalized marketing system from architecture through production deployment, supported by empirical results from a 24-month e-commerce dataset. The proposed hybrid model—combining ML-based customer clustering, BERT-based NLP inference, deep learning recommendation, and contextual bandit re-ranking—achieves classification accuracy in the 93–95% range and satisfies response latency and retention targets in the production environment. The performance improvements over the rule-based baseline are substantial and consistent across evaluation runs.

The paper has attempted to report limitations with the same prominence as results. The dataset is drawn from a single platform operating in a specific market, and the reported metrics should not be assumed to transfer without validation to other contexts. The CLV metric is subject to causal confounding. The differential privacy budget is permissive and may not satisfy all regulatory interpretations. The NLP pipeline underperforms on Hindi and Marathi inputs relative to English. Each of these limitations defines a concrete direction for future work.

The Model Staleness Index, documented here for the first time, provides a practical mechanism for triggering retraining based on distributional drift before accuracy degradation becomes visible in business metrics. Its calibration methodology and sensitivity characteristics warrant further investigation across different platforms and behavioural contexts. The broader message of this paper is that deploying AI in marketing at production scale requires equal attention to data engineering, system reliability, privacy compliance, and model accuracy—and that honest documentation of failures and limitations is as valuable to the research community as documentation of successes.

ACKNOWLEDGMENTS

The authors thank the JSPM University Research Ethics Committee for expeditious review (reference: JSPM/REC/2022/114) and the platform operator for providing data access under a signed data processing agreement. The authors disclose the use of an AI writing assistant (Claude, Anthropic) for initial structural drafting; all technical content, experimental design, results interpretation, and editorial judgements are the authors' own. No external funding was received for this research.

REFERENCES

- [1] P. Domingos, "A few useful things to know about machine learning," *Commun. ACM*, vol. 55, no. 10, pp. 78–87, Oct. 2012.
- [2] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 6, pp. 734–749, Jun. 2005.
- [3] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. Conf. North American Chapter Assoc. Comput. Linguistics: Human Language Technol. (NAACL-HLT)*, Minneapolis, MN, USA, Jun. 2019, pp. 4171–4186.
- [4] T. Davenport, A. Guha, D. Grewal, and T. Bressgott, "How artificial intelligence will change the future of marketing," *J. Acad. Marketing Sci.*, vol. 48, no. 1, pp. 24–42, Jan. 2020.
- [5] V. Kumar and W. Reinartz, *Customer Relationship Management: Concept, Strategy, and Tools*, 3rd ed. Berlin, Germany: Springer, 2018.
- [6] European Parliament and Council of the EU, "Regulation (EU) 2016/679 of the European Parliament and of the Council on the protection of natural persons with regard to the processing of personal data (General Data Protection Regulation)," *Official Journal of the European Union*, vol. L 119, pp. 1–88, Apr. 2016.
- [7] L. Chen, P. Goes, and A. Marsden, "The ethics of AI in marketing: Fairness, transparency, and accountability," *IEEE Trans. Technol. Soc.*, vol. 2, no. 4, pp. 150–160, Dec. 2021.
- [8] P. C. Verhoef, K. N. Lemon, A. Parasuraman, A. Roggeveen, M. Tsiros, and L. A. Schlesinger, "Customer experience creation: Determinants, dynamics, and management strategies," *J. Retailing*, vol. 85, no. 1, pp. 31–41, Mar. 2009.
- [9] C. Molnar, *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*, 2nd ed. Munich, Germany: lulu.com, 2022. [Online]. Available: <https://christophm.github.io/interpretable-ml-book/>
- [10] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Nov. 2011.
- [11] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.
- [12] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006.
- [13] F. Ricci, L. Rokach, and B. Shapira, Eds., *Recommender Systems Handbook*, 2nd ed. New York, NY, USA: Springer, 2015.
- [14] J. B. Schafer, J. A. Konstan, and J. Riedl, "E-commerce recommendation applications," *Data Min. Knowl. Discov.*, vol. 5, no. 1–2, pp. 115–153, Jan. 2001.
- [15] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001.
- [16] G. Adomavicius and A. Tuzhilin, "Context-aware recommender systems," in *Recommender Systems Handbook*, 2nd ed., F. Ricci, L. Rokach, and B. Shapira, Eds. New York, NY, USA: Springer, 2015, pp. 191–226.
- [17] R. Sennrich, B. Haddow, and A. Birch, "Neural machine translation of rare words with subword units," in *Proc. 54th Annu. Meeting Assoc. Comput. Linguistics (ACL)*, Berlin, Germany, Aug. 2016, pp. 1715–1725.
- [18] M. Abadi et al., "Deep learning with differential privacy," in *Proc. ACM SIGSAC Conf. Comput. Commun. Security (CCS)*, Vienna, Austria, Oct. 2016, pp. 308–318.

- [19] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated machine learning: Concept and applications," ACM Trans. Intell. Syst. Technol., vol. 10, no. 2, Art. no. 12, Feb. 2019.
- [20] Ministry of Electronics and Information Technology, Government of India, "The Digital Personal Data Protection Act, 2023," Gazette of India, New Delhi, India, Aug. 2023. [Online]. Available: <https://www.meity.gov.in/dpdpa>
- [21] L. Li, W. Chu, J. Langford, and R. E. Schapire, "A contextual-bandit approach to personalized news article recommendation," in Proc. 19th Int. Conf. World Wide Web (WWW), Raleigh, NC, USA, Apr. 2010, pp. 661–670.
- [22] M. Beauchemin et al., "Apache Airflow: A workflow management platform for data engineering pipelines," Apache Software Foundation, 2015. [Online]. Available: <https://airflow.apache.org>



Pranjali Dhanbhar is a master's student in the Department of Computer Applications at JSPM University, Pune, Maharashtra, India. Her research interests include artificial intelligence, machine learning systems, and data privacy in enterprise applications.



Kalpana Salunke

is a faculty member in the Department of Computer Applications at JSPM University, Pune, Maharashtra, India. Her research interests include applied machine learning, database systems, and software engineering methodologies.