Contextual Bandits for Personalized Subscription Plan Offerings in Digital Content Services

Luigi Pio Battista*, Saverio Ieva*†, Filippo Gramegna*†,
Floriano Scioscia*†, Michele Ruta*†, Alessio Robertazzi‡, Rocco Michele Lancellotti§,
* Polytechnic University of Bari – Via E. Orabona 4, Bari (I-70125), Italy – {name.surname}@poliba.it
† donkeyPower S.r.l. – Via E. Orabona 4, Bari (I-70125), Italy – {name.surname}@donkeypower.it
‡ Go Reply S.r.l. – Via Robert Koch 1, Milan, Italy – g.robertazzi@reply.it
§ RCS MediaGroup S.p.A. – Via Rizzoli 8, Milan, Italy – michele.lancellotti@rcs.it

***Abstract*—Subscription-based digital content services are progressively adopting adaptive strategies to personalize offerings and enhance user engagement. This study proposes a contextual bandit framework for the dynamic selection of personalized subscription plans, leveraging user behavior profiles and contextual features to generate tailored recommendations. The framework has been evaluated in a real-world deployment in collaboration with *RCS Innovation S.r.l.*, where it supports call center operators in identifying the most suitable subscription options for individual users. Early experimental outcomes demonstrate high predictive accuracy and support the applicability of the approach in production environments.**

***Index Terms*—Multi-armed bandits, contextual bandits, subscription personalization, revenue maximization, user behavior modeling**

I. Introduction

In the context of the evolving knowledge society, smart communities are increasingly characterized by pervasive access to multimedia content, including text, audio, and video, all delivered in digital formats. Subscription-based offerings have become the predominant content monetization model, providing convenience for users and predictable revenue streams for service providers. To maintain a competitive advantage, content producers and aggregators are required to implement sophisticated strategies that support the personalization of both content offerings and pricing or subscription modalities. This requirement has led to a growing interest in Machine Learning (ML) techniques that exploit user data and behavioral feedback to dynamically adapt subscription plans and promotional strategies. Among the various approaches, *Multi-Armed Bandits (MAB)* [1] represent a well-established framework for sequential decision-making problems under uncertainty. The MAB paradigm models the interaction between a decision-making agent (or policy) and a stochastic environment as a probabilistic process. In this framework, each possible action, referred to as an "arm", produces a reward drawn from an unknown probability distribution. The decision policy iteratively selects actions and refines its strategy based on the rewards received. This iterative process involves a trade-off between exploration, aimed at acquiring new information, and exploitation, focused on leveraging known information to maximize rewards.

Classical policies, including *Epsilon-Greedy*, *Upper Confidence Bound*, and *Thompson Sampling* [2], perform well under stationary reward assumptions, but are inadequate in personalized digital services where user-specific factors – such as demographics, behavior, and interaction history – strongly affect outcomes. To address this need, the *Contextual Bandits (CB)* framework [1] extends the classical MAB model by integrating context available at each decision point. For example, in [3] a CB-based policy is used to sequentially select articles for presentation based on user-specific context and to update its strategy over time using click-based feedback, with the objective of maximizing overall user engagement. Complementary methodologies have also been explored. For instance, Kao et al. [4] propose a Bayesian framework aimed at optimizing subscription offers and durations under uncertainty in user behavior and service costs. Although this approach does not employ CB models, it similarly addresses early-stage uncertainty by employing Beta prior distributions and smoothing techniques, which serve to enhance the robustness of initial decisions prior to the availability of sufficient empirical evidence.

This paper introduces a CB-based framework for personalized subscription plan selection, aligning recommendations with individual user profiles. It supports three operational modes – standard, retention, and revenue – allowing marketing teams to prioritize different strategic goals. At inference time, the system generates up to three tailored proposals per user, each optimized according to the selected mode. The structure of the paper is as follows: Section II details the proposed framework, Section III describes a real-world case study alongside a preliminary evaluation, and final remarks are in Section IV.

II. Proposed Framework

Figure 1 summarizes the CB-based framework, which personalizes subscription plans through a closed-loop interaction among predictive models, policy, and reward driven by user context.

Context ingestion. In each iteration, the framework ingests contextual inputs (e.g., demographic attributes, behavioral patterns), which are processed into a fixed-length feature vector. This context representation is fed to a collection of

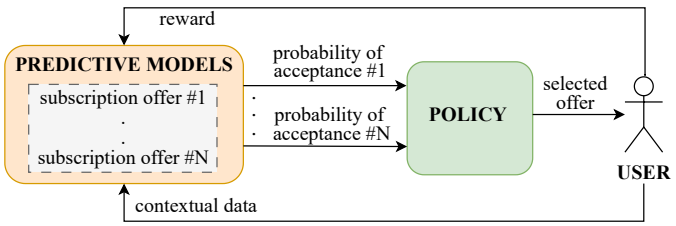

Fig. 1. Framework architecture and training workflow.

offer-specific predictive models, each of which estimates the acceptance probability for its respective subscription offer.

Probability estimation. Each predictive model acts as a supervised learning oracle, estimating the likelihood of user acceptance based on contextual features. This component is model-agnostic and supports classifiers capable of probabilistic outputs, such as logistic regression, decision trees, gradient boosting, or neural networks. The proposed implementation adopts an XGBoost classifier for its strong performance and interpretability on structured data.

Policy selection. The resulting probabilistic estimates inform the policy component, which select a single subscription offer to present to the user. The policy is designed to manage the trade-off between exploration and exploitation, and supports multiple selection strategies (e.g., Bootstrapped Upper Confidence Bound, SoftmaxExplorer), allowing for flexible adaptation based on empirical validation [5].

Reward feedback. The user's response to the selected offer generates a reward signal, represented as a binary outcome indicating acceptance or rejection. This reward serves a dual purpose: it informs the policy for subsequent decision-making and updates the training history of the predictive model associated with the selected offer. Through this feedback mechanism, the system incrementally refines its decision strategy with the objective of maximizing the expected cumulative reward over time.

To mitigate the cold-start problem, the framework integrates smoothing techniques and employs Beta prior distributions to regularize early-stage estimates for offers with limited interaction data. This approach aims to enhance the reliability of initial decisions in the absence of sufficient historical feedback.

III. Early Evaluation

A practical case study involving *RCS INNOVATION S.R.L.* has assessed the framework within a call center environment, where it aids operators by suggesting up to three customized subscription plans for each user. The action space includes four subscription types in monthly and annual formats, organized into pricing tiers. To balance complexity and relevance, a curated set of representative offers has been modeled, each as a distinct arm within the contextual bandit framework. The system has been trained using user-specific contextual features, which include a three-month browsing history, user's device and geographic information, as well as the current subscription status at the time of the call. The reward signal is binary: in the *standard* operational mode, a reward of 1 is assigned when the model-selected offer aligns with the historical proposal made by the call center. Two alternative reward configurations have been designed to reflect specific marketing objectives: (i) *retention mode*, which emphasizes offers associated with lower churn rates, and (ii) *revenue mode*, which prioritizes options with higher estimated LifeTime Value (LTV). Churn probabilites and LTV metrics are computed from historical subscription data, capturing one-year attrition probability and expected revenue, respectively. Model training is performed monthly, while inference is executed daily to generate three personalized proposals per user across all operational modes. Final model selection is guided by *top-3* accuracy, measured at the final iteration, with consistent performance ranging between 0.85 and 0.9 on the test set. The complete pipeline has been deployed on Google Cloud Platform (GCP), using Kubeflow Pipelines and Dataproc Serverless for data preprocessing. Workflow orchestration is managed through Vertex AI Pipelines, while Apache Airflow is employed for task scheduling, metric logging and alerting.

IV. Conclusion

This work has proposed a CB-based framework for the adaptive personalization of subscription offers, aimed at aligning pricing strategies with individual user profiles and diverse marketing objectives. By leveraging contextual information and supporting multiple reward configurations, the framework simplifies data-driven optimization across distinct operational modes. Preliminary results demonstrate promising performance, particularly in terms of *top-3* accuracy and suitability for real-time deployment. The framework has been successfully integrated into a production-level call center workflow. Future research directions include incorporating feedback from human operators and investigating online learning techniques to further enhance the framework's adaptability and long-term effectiveness.

Acknowledgments

This work has been supported by the *RCS Digital* grant (code EVY3OY8), co-funded by RCS Innovation S.r.l. and European Regional Development Fund for Apulia Region 2014/2020 Operating Program.

References

[1] A. Slivkins, *Introduction to Multi-Armed Bandits*. Now Foundations and Trends, 2019.

[2] G. Elena, K. Milos, and I. Eugene, "Survey of multiarmed bandit algorithms applied to recommendation systems," *International Journal of Open Information Technologies*, vol. 9, no. 4, pp. 12–27, 2021.

[3] L. Li, W. Chu, J. Langford, and R. E. Schapire, "A contextual-bandit approach to personalized news article recommendation," in *Proceedings of the 19th International Conference on World Wide Web*, 2010, pp. 661–670.

[4] Y.-M. Kao, N. B. Keskin, and K. Shang, "Bayesian dynamic pricing and subscription period selection with unknown customer utility," *SSRN preprint 3722376*, 2024.

[5] D. Cortes, "Adapting multi-armed bandits policies to contextual bandits scenarios," *arXiv preprint arXiv:1811.04383*, 2018.