ENCODE: Ensemble Contextual Bandits in Big Data Settings - A Case Study in E-Commerce Dynamic Pricing

Srividhya Sethuraman
TCS Research
Tata Consultancy Services
Chennai, India
srividhya.sethuraman1@tcs.com

Uma Maheswari G
TCS Research
Tata Consultancy Services
Chennai, India
uma.mg@tcs.com

Siddhesh Thombre
TCS Research
Tata consultancy Services
Chennai, India
siddhesh.thombre@tcs.com

Sunny Kumar
TCS Research
Tata Consultancy Services
Chennai, India
sunny.k10@tcs.com

Vikash Patel
TCS Research
Tata Consultancy Services
Chennai, India
patel.vikash5@tcs.com

Dr. Sharadha Ramanan
TCS Research
Tata Consultancy Services
Chennai, India
sharadha.ramanan@tcs.com

2023 IEEE International Conference on Big Data (BigData) | 979-8-3503-2445-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/BigData59044.2023.10386412

***Abstract*—We present ENCODE, an innovative ensemble-based Contextual Bandit (CB) model, engineered explicitly for dynamic pricing in large-scale e-commerce for kid's clothing. ENCODE uniquely addresses the retailer's multifaceted need for optimal pricing — both immediate and cumulative — subject to market-driven price triggers like competitor fluctuations and seasonal trends. The model integrates four cornerstone CB algorithms: LinUCB, Vowpal Wabbit, Contextual Thompson Sampling, and BayesUCB, delivering a unified solution for maximizing yield in alignment with business constraints.**

Our model's originality manifests in its comprehensive treatment of complex, real-world pricing issues: from optimizing long-tail products with limited data to seamlessly managing an extensive array of products and catering to diverse customer segments. ENCODE also uniquely accommodates pricing strategies within product families, is highly responsive to shifts in competitor pricing, and accounts for the intricate interplay between related products in a non-stationary environment.

Demonstrably effective, the model yielded a 13.8% improvement in margin through Contextual Thompson Sampling alone, with an additional 6% gain from considering inter-related products. In comparative analysis, ENCODE surpassed standalone CB algorithms by 19% in cumulative margins. It also boasts computational efficiency, optimized through dimensionality reduction and hyper-parameter fine-tuning. Scalability is ensured through advanced cloud-based implementations. Hence, ENCODE stands as a comprehensive, highly scalable, and markedly effective solution to contemporary e-commerce pricing complexities.

***Index Terms*—Contextual Bandits, Multi-Armed Bandits, Reinforcement learning, Online Learning, E-commerce, LinUCB, BayesUCB, Vowpal Wabbit, Thompson Sampling, Ensemble model, Long tail Products, Scaling, Big Data, Bayesian Model**

I. Introduction

The e-commerce sector is characterized by a fluid and rapidly evolving environment, where pricing strategies are central to a retailer's success. Working with a major kids' clothing e-commerce retailer — referred to as Dataset 1 — we have identified a myriad of intricate, yet previously unexplored, challenges in this space. Among these is the need for a pricing model capable of providing price recommendations so that the margin is maximized not only for the immediate moment but also for a continuous period until a market-driven price trigger necessitates change.

To address these nuanced complexities, we introduce ENCODE (ENsemble of Contextual bandits for Dynamic pricing in E-commerce), a groundbreaking ensemble-based Contextual Bandit (CB) model. ENCODE amalgamates the strengths of four key CB algorithms: LinUCB, Vowpal Wabbit, Contextual Thompson Sampling, and Bayes UCB. Uniquely, our model goes a step further by incorporating an ensemble level to critically review immediate pricing suggestions for their long-term applicability, thereby balancing margin stability and customer satisfaction.

Our work contributes to the field in three distinct ways. First, we employ ensemble learning to integrate various CB algorithms, fortifying the robustness and diversity of our pricing recommendations. This enables ENCODE to cater to a wide array of product categories, such as regular, seasonal, and fashion items, and outperforms individual CB models in terms of cumulative margins.

Second, we subject ENCODE to rigorous real-world validations. The model is calibrated to respond dynamically to a myriad of variables like competitor pricing, customer segmentation, and time-sensitive factors including seasonality and economic shifts. We validate ENCODE's efficacy across three disparate datasets, each with its own challenges, and find that it consistently delivers robust, optimal solutions while achieving key performance metrics such as revenue lift, margin

enhancement, and effective inventory clearance.
Third, ENCODE is architected for scalability, aptly demonstrated in our case studies involving a vast number of products. We combat the challenges associated with high dimensionality through techniques like dimensionality reduction and hyperparameter tuning. Additionally, we employ advanced methods such as approximate inference, caching, model approximation, and sampling reductions to ensure computational efficiency. The model is further scaled via state-of-the-art cloud technologies and distributed computing frameworks for parallel execution.
In conclusion, ENCODE is a novel ensemble-based CB model that comprehensively and effectively tackles the multifaceted challenges inherent in modern e-commerce pricing, offering a robust and highly scalable solution.

II. Related work

In recent years, dynamic pricing in e-commerce has garnered significant attention within research circles. Existing studies approach this subject from various angles, each contributing to the understanding of pricing mechanisms. In [1], the focus is on understanding the unknown demand curve by treating buyer valuations as identical, random, and worst-case scenarios. Diverging from this, our research leverages covariates as side information, utilizing a parametric demand model to extract more meaningful insights. Similarly, an extensive survey in [2] delineates dynamic pricing under two key paradigms: 1) fluctuating demand functions, and 2) static demand influenced by inventory levels. We extend this by demonstrating how ensemble models can offer superior rewards compared to individual bandit policies, especially in contexts that align with both of these settings. In our ENCODE model, the CB algorithms actively adapt to these changing dynamics by assimilating contextual features and continuously evolving their price recommendations. Another noteworthy contribution is from [3], which investigates dynamic pricing for products with high-dimensional features, albeit without inventory constraints in a theoretical setting under different assumptions. In contrast, our work is set in a real-world e-commerce environment and recommends practical modifications that optimize computing resources associated with its deployability. This differentiates us from other works on bandits with side-information such as those in [4] [5] [6] [7], where free exploration is often employed. In our model, the set of arms is finite and tied to a distinct set of prices, thus providing no additional information about the rewards of other arms when one is selected. Furthermore, the study in [5] proposes ellipsoid-based dynamic pricing for highly differentiated products, assuming that product features predominantly drive market values. In contrast, our model accounts for both product features and additional temporal factors that uniquely influence sales. Lastly, whereas studies [8] [9] [10] [11] explore the applicability of bandit policies in recommendation systems, our work is novel in its focus on employing an ensemble bandit specifically for dynamic pricing. In summary, our work presents an innovative blend of ensemble modeling and CB algorithms, designed to address the intricate dynamics of real-world e-commerce pricing.

III. Data characteristics

In our study, we employ three distinct but complementary datasets to deepen the understanding of dynamic pricing in the e-commerce space. Our primary dataset, referred to as Dataset 1, is synthetically generated based on models [12] derived from a large kids' clothing e-commerce retailer. This dataset is meticulously structured to encompass various metrics, including performance, attributes, pricing, and inventory across 14 categories. It features 1,905 styles and 9,257 products, focusing particularly on styles associated with the Fall season, from August to late February. These styles, comprising over 500 articles each, are grouped under the same merchandise hierarchy — like color-pattern combinations. Throughout our study, terms such as 'articles,' 'products,' 'SKUs', and 'items' are used interchangeably, underscoring their conceptual similarity.
Our second dataset originates from H&M and is publicly available on Kaggle [1]. It encompasses transactional, product attribute, and customer data, providing a comprehensive look at online retail operations.
The third dataset, drawn from the same H&M source but specifically spotlighting Fashion SKUs, complements our primary dataset. Key statistics from all three datasets are summarized in Table I. What sets our work apart is the nuanced approach to competitive pricing strategy, evidenced in Figures 1 and 2. Here, the retailer's prices are juxtaposed against those of five active competitors for a high-selling style from Dataset 1. As Figure 2 elaborates, the retailer doesn't adhere to a fixed strategy when responding to competitors, allowing for a flexible, time-lagged reaction to market conditions.
Another innovative aspect is our focus on customer segmentation in Dataset 3, which reveals year-over-year buying patterns within certain customer groups. These are chiefly loyal customers, differentiated mainly by age. Figure 3 plots the first two principal components, illustrating the variances between these customer clusters. This nuanced understanding of customer behavior adds another layer of depth to our analysis, making it both original and robust.

TABLE I: Summary of key statistics for the datasets

Dataset	Train period (Year-Week)	Test period (Year-Week)	Price (in USD)	Sales units
Dataset 1	201936 - 201949	202036 - 202049	4.7 - 62.7	2 - 1100
Dataset 2	201944 - 202005	201801 - 201835	0.75 - 49.99	1 - 676
Dataset 3	201918 - 201927	202018 - 202027	6 - 35	1 - 543

IV. Methodology

Our methodology deploys an advanced framework for dynamic pricing, uniquely leveraging robust data analytics and

[1]https://www.kaggle.com/competitions/h-and-m-personalized-fashion-recommendations/data

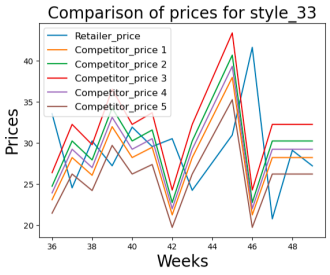

Fig. 1: Retailer prices versus prices of five active competitors

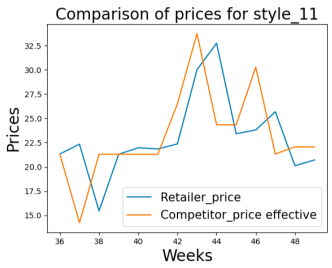

Fig. 2: Retailer prices versus prices of effective competitor

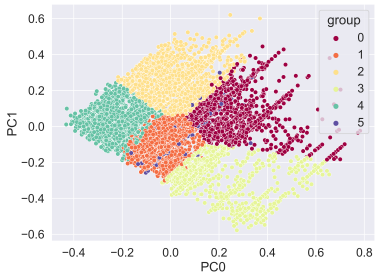

Fig. 3: Visualization of first two principal components of customer clusters

Machine Learning (ML), as depicted in Figure 4. We integrate diverse data streams — transactional, attribute-based, inventory, and price master data — which not only reflect historical sales but also fuel ML-based demand forecasts, chiefly via XGBoost. A novel element is the application of Thompson Sampling-based models for calculating price elasticity, confirming its efficacy through our previous work [13]. Our data inputs are further enriched with item-specific attributes, pricing details, temporal markers, and discount indicators. They are further refined by high-frequency market data to capture real-time dynamics in the e-commerce sphere. This consolidated dataset undergoes rigorous cleaning and preprocessing, including feature extraction and dimensionality reduction, before clustering and aggregation to suit business goals. Upon processing, a Contextual Bandit (CB) algorithm is activated, steered by predefined triggers such as competitor pricing and market trends. An ensemble model then oversees price recommendations, ingeniously employing epsilon-greedy and Q-learning strategies to focus on immediate and future gains, thereby securing optimal pricing. This ensemble is regulated by a critical hyperparameter — batch size — which aligns our model with the retailer's dual focus on immediate results and long-term stability, typically for 1–2 weeks. Projections generated by an array of ML and Deep Learning models grant this forward-looking perspective. Further distinctiveness is manifested in our ENCODE model, which employs four CB algorithms: LinUCB, Vowpal Wabbit (VW) mini-monster, Contextual Thompson Sampling (CTS), and Bayes UCB. These algorithms are benchmarked against the

Algorithm 1 ENCODE model pseudocode using Q-Learning

Initialize hyperparameters: α, Γ, $num\_episodes$
Initialize Q-values for each product and price combination: q_values
Initialize a dictionary to record the chosen bandit policies: $bandit\_policy\_counts$
while $num\_episodes > 0$ **do**
  for $i \leftarrow 1$ to $num\_products$ **do**
    $context\_vector \leftarrow$ generate_context(i)
    $policy\_idx \leftarrow$ argmax($q\_values[i]$)
    $sel\_policy \leftarrow$ bandit_policies[$policy\_idx$]
    Record the chosen bandit policy as sel_policy
    $prices \leftarrow sel\_policy(context\_vector, i)$
    $price\_index \leftarrow$ argmax(reward) for $p \in prices$
    $pid \leftarrow price\_index$
    $optimal\_price \leftarrow prices[pid]$
    $Sales\_projection \leftarrow sales\_projection$(future)
    $q\_values[i, pid] \leftarrow q\_values[i, pid] + \alpha \cdot (reward + \Gamma \cdot \max(q\_values[i]) - q\_values[i, pid])$
  end for
  $num\_episodes \leftarrow num\_episodes - 1$
end while

Classical Non-linear pricing optimization using a Sequential Least SQuares Programming (SLSQP) baseline and consider multifaceted context features to optimize a composite reward function balancing sales, revenue, and margin. Our ENCODE model specializes in amalgamating various CB algorithms to produce price recommendations, underpinned by a cumulative reward metric as expressed in Equation 7. The psuedocode for our ensemble model is given in Algorithm 1 wherein a Q-learning algorithm is employed in level 1 which oversees the price recommendations of the CB algorithms in level 0. Here, Γ, where $\Gamma < 1$, is a hyperparameter that weighs immediate rewards over future rewards and can be suitably tweaked as per the retailer requirements. Uniquely, we introduce a second level of bandit algorithms to supervise these recommendations. We also offer the flexibility to employ any Reinforcement Learning policy as an alternative. In summary, our ENCODE model, showcased in Figure 4, ensures pricing decisions that are not just timely but also future-proof, based on real-time and forecasted data. This framework uniquely accommodates "what-if" scenarios for performance evaluation and aligns with margin-focused strategies as outlined in section (3) of section VI. It takes into account not just the current market context but also projects future trends, thereby enabling retailers to make pricing decisions that are optimized for both immediate and extended timeframes. In doing so, we offer an innovative, data-driven, and comprehensive solution for dynamic e-commerce pricing.

A. Contextual bandits based Online Dynamic Pricing

The inputs to the ENCODE Model include duration of the selling horizon and weights for the objective function. The context features comprise product attributes, price

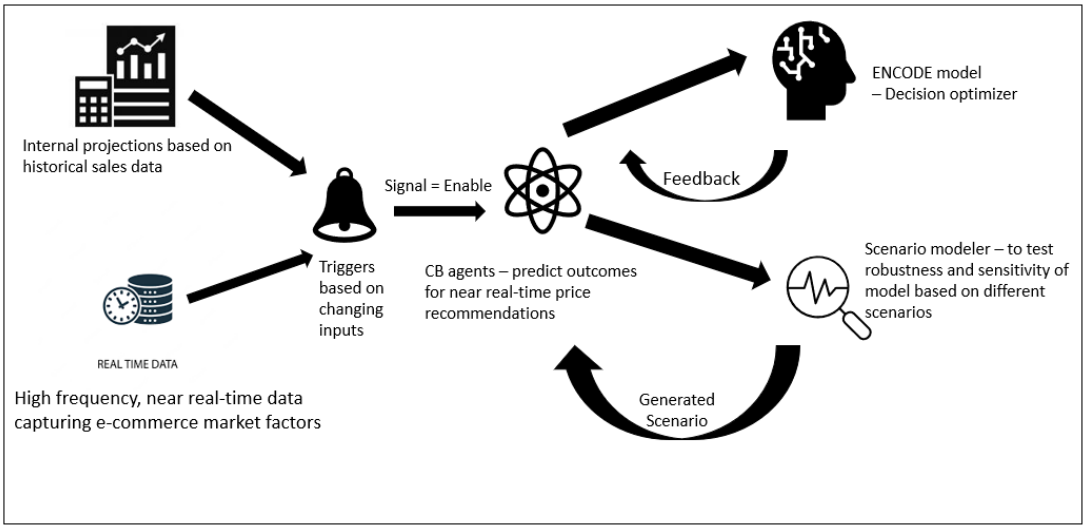

Fig. 4: ENCODE system architecture

attributes, temporal attributes, and customer attributes. The action constitutes discrete price points within the minimum and maximum price range, and the reward is a weighted combination of sales, revenue and margin. The objective of the CB algorithms is to maximize the average reward. The four CB algorithms applied for dynamic pricing in our setting are: LinUCB, VW mini-monster algorithm, CTS and Bayes UCB. For comparison with passive classical optimization algorithms, SLSQP is used.

LinUCB Algorithm assumes linear dependency between the expected reward of an action and its context and models the representation space using a set of linear predictors. Contextual Thompson Sampling uses a heuristic for choosing action for a context, that maximizes the expected reward with respect to a randomly drawn belief. Prior is assumed for distribution parameters for probability distribution of action and likelihood is based on reward given action, and context. Bayes UCB is a general formulation for a class of Bayesian index policies that relies on quantiles of the posterior distribution. In this paper, we employed different VW CB algorithms like Epsilon Greedy (default exploration), Explore First, Bagging Explorer, Online Cover, and SoftMax Explorer and compared their results.

Context : The context features consist of - Product attributes, Price attributes, Temporal attributes, and Customer attributes.

Action : Discrete price points within minimum, maximum price range

Reward : Weighted sum of sales, revenue and margin.

B. Dynamic Pricing Formulation

We study the scenario in which an e-commerce retailer sells a set of non-perishable products with unlimited availability. Given a discrete time horizon of T rounds, in each round t, the policy (representing the 'Online retailer') selects a price P_t within specified minimum and maximum price bounds. The demand D_t in this round is then independently drawn from a fixed distribution with unknown parameters. The average reward of the CB algorithm at the end of the selling horizon for 'N' iterations and 'K' styles is given in Equation 1. Dynamic pricing can be modelled as a Sequential Decision Process since price in the next round does not depend on the previous rounds, given the current price and price bounds. Hence, the CB algorithms are applicable for this problem.

The objective function of the CB algorithm for pricing is to maximize the total cumulative reward for the styles. The reward for a style is calculated as the yield which is the weighted sum of sales, revenue and margin and is given by Equation 2. The pricing policy must satisfy the constraint that the prices recommended for a style must lie between the minimum and maximum price for the style as specified by the retailer, as shown in Equation 6. The demand model given by Equation 3, considers price elasticity, inter-related item effects and competitor effect. Refer Table II for the descriptions of notations.

$$\text{Average reward} = \sum_{\text{iteration = 1}}^{N} \sum_{\text{style = 1}}^{K} \frac{\text{Rewards for each style}}{\text{No: of iterations, N}} \tag{1}$$

$$Maximize \;\; Yield = \alpha_1 * Sales + \alpha_2 * Revenue + \alpha_3 * Margin \tag{2}$$

$$S_{new,i}[t] = S_{pred,i}[t] * \frac{(\frac{p_{opt,i}[t]}{p_{init,i}[t]})^{-\gamma_{i,j}}}{exp(\frac{lr_i * (P_{opt,i}[t] - p_{comp,i}[t]}{p_{comp,i}[t]})} \tag{3}$$

$$Revenue = S_{new,i}[t] * P_{opt,i}[t] \tag{4}$$

$$Margin = S_{new,i}[t] * (P_{opt,i}[t] - cp_i) \tag{5}$$

such that,

$$P_{min,i} \leq P_{opt,i}[t] \leq P_{max,i} \tag{6}$$

$$Cumulative\ \ reward = Yield\ \ + \sum_{n=1}^{T} \Gamma^n * reward(t+n) \tag{7}$$

where, $i \in$ style, $j \in$ inter-related style.

TABLE II: Table of Notations

Notation	Description
$P_{init,i}$	initial/previously updated price forecast style 'i'
$S_{pred,i}[t]$	Sales forecast for the price P_{init} for round 't' for style 'i',
$\gamma_{i,j}$	cross-price elasticity between style 'i' & style 'j'
$P_{opt,i}[t]$	optimal price for the round 't' for style 'i'
$S_{new,i}[t]$	sales corresponding to optimal price for round 't' for the style 'i', considering inter-item & competitor effects,
cp_i	unit cost price of style 'i'
$\alpha_1, \alpha_2, \alpha_3$	weights for the objective, i.e.,sales, revenue & margin such that $\alpha_1 + \alpha_2 + \alpha_3 = 1$,
$P_{min,i}, P_{max,i}$	min-max bounds for price of style 'i'
lr_i	competitor leakage coefficient for style 'i'
$P_{comp,i}[t]$	competitor price for round 't' for style 'i'

V. Experiments and Insights

In this section, we showcase the efficacy of the proposed ENCODE model and discuss relevant insights. We also showcase the performance of CB algorithms for different applications in retail. For simulation, we pick random instances corresponding to the three datasets. The objective was set to margin maximization (setting α_1 and α_2 as 0 in Equation 2) with the set of constraints discussed in the previous section.

Our comprehensive results are summarized in Table III. A notable discovery from Table III is that the Contextual Thompson Sampling algorithm outperforms others on Dataset 1, delivering an impressive 13.8% improvement in margin. For Dataset 2, Vowpal Wabbit algorithms excel, resulting in a 7% margin boost. Meanwhile, the CTS algorithm also stands out in Dataset 3, showing a 5% margin enhancement. Cumulatively, our CB algorithms demonstrate a remarkable 30% higher reward compared to traditional passive learning approaches. This variability in rewards across different CB algorithms underlines the merit of our ensemble strategy, ingeniously designed to exploit the unique advantages of each individual algorithm.

1) Validation of our CB algorithms: To authenticate the performance of our CB algorithms, we employ two distinct validation approaches. First, for Dataset 1, we rely on the computation of average reward, which consistently improves over iterations, as demonstrated in Figure 5. In the case of Datasets 2 and 3, we introduce an innovative Oracle mechanism. The Oracle is calibrated through numerous iterations to determine the maximum achievable reward in each context. We then compute regret as the difference between this optimal reward and the reward obtained from our specific pricing solutions. Notably, our method consistently yields lower regret values, moving us closer to the global optimum, as corroborated in Figure 6. By continually tracking the average reward and regret across iterations, we are able to conclusively validate that our CB algorithms are exceptionally effective in achieving the targeted objectives. Our methodology therefore not only presents state-of-the-art solutions for dynamic pricing but also contributes to the literature by providing robust and verifiable validation mechanisms.

Algorithm 2 Tweaked CTS for Dynamic Pricing

Initialize $num\_products$, $num\_features$, $price\_recommendations$, $mean\_posterior$, $covariance\_posterior$, arm_pulls, $arm\_rewards$, $cache$
function UpdatePosterior($product$, $context$, $reward$)
    Update $mean\_posterior$ and $covariance\_posterior$ using Gaussian posterior equations
end function
function SampleFromPosterior($product$)
    Obtain $sample$ from Gaussian distribution with $mean\_posterior$ and $covariance\_posterior$
    return $sample$
end function
function ThompsonSampling($context$)
    for $prod = 1$ to $num\_products$ **do**
        if $prod$ not in $cache$ **then**
            Cache $sample$ from Posterior($prod$)
        end if
        $sampled\_rewards[prod] = context \cdot sample$
    end for
    $chosen\_product = \arg\max(sampled\_rewards)$
    Implement dynamic pricing logic for $chosen\_product$
    $reward = calculate\_reward(chosen\_product)$
    $arm\_pulls[chosen\_product] += 1$
    $arm\_rewards[chosen\_product] += reward$
    return $price\_recommendations[chosen\_product]$
end function

TABLE III: Cumulative margin from different CB algorithms for different datasets

Datasets	Average reward CTS	Average reward VW	Average reward LinUCB	Average reward BayesUCB	Average reward SLSQP	Average reward ϵ-greedy
Dataset 1	**40998.5**	39567.5	32345.81	32238.45	35998	37431.7
Dataset 2	47865.67	**48544.35**	34138.15	39618.15	32241	40326.5
Dataset 3	**385294.0**	379348.2	327542.6	327811.0	298650	31659.4

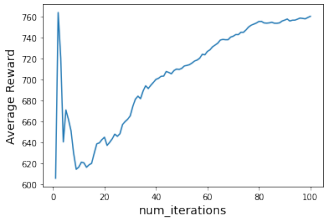

Fig. 5: Average reward across iterations - Dataset 1

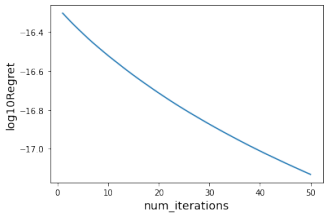

Fig. 6: Plot of Regret across iterations - Dataset 2

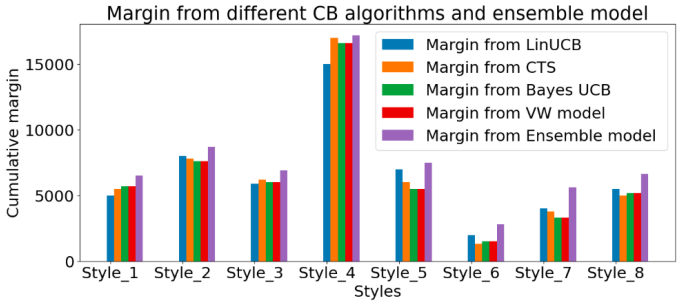

Fig. 7: Cumulative margin for different CB algorithms and the ENCODE model

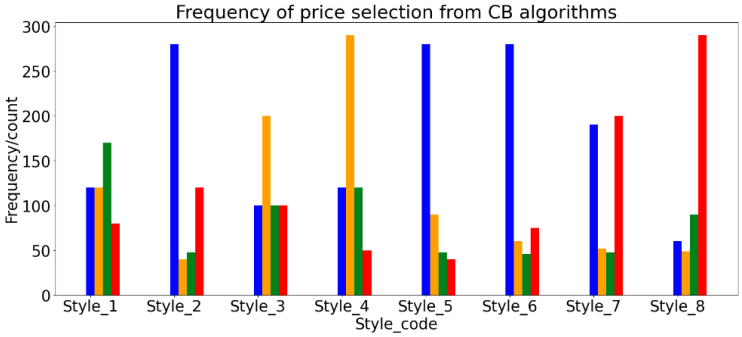

Fig. 8: Frequency of price selection from different CB algorithms by ENCODE

ENCODE Versus Individual models In this section, we underscore the innovative superiority of our ensemble ENCODE model over individual models in the context of dynamic pricing, particularly using Dataset 1 as a case study. The ENCODE model not only augments accuracy and performance but also mitigates risks associated with over-fitting and under-fitting by judiciously balancing bias-variance trade-offs. For new products, it is challenging to come up with optimal prices as these products do not have sufficient historical sales data. For example, consider a new electronic gadget which has an enhanced set of features than all its predecessors. In such cases, these products do not have very close substitutes whose selling characteristics could be leveraged. This is often referred to as one of the cold-start problems in pricing literature. Our ENCODE model incorporates exploration of different price points both in the individual bandit algorithms as well as in the ensemble model and thus, ingeniously addresses the cold start problem associated with pricing new products without historical sales data.

Our empirical results compellingly validate the efficacy of the ensemble approach. Figure 7 reveals that the ensemble model significantly outpaces individual bandit algorithms in achieving higher cumulative margins. Specifically, for a subset of eight representative styles, the ensemble model delivers a 19% improvement in cumulative margin over individual models. This is further illustrated in Figure 8, which depicts the frequency of price selection by the ensemble model across these styles. In Figure 8, the blue bar represents prices from LinUCB algorithm, orange, green and red bars represent price recommendations from CTS, BayesUCB and VW algorithms respectively. This way, our ensemble bandit model helped us replicate A/B testing [8] of price recommendations in e-commerce. ENCODE decides between price recommendations offered by different CB algorithms. If one of them is clearly less effective than the others, it will progressively reduce the number of times that price recommendation gets adopted. Table IV provides additional granularity, showing the frequency of price selections from the different CB algorithms and their average cumulative margins respectively, thus solidifying the ensemble model's dominance in our framework.

In summary, our novel ensemble model - ENCODE, not only significantly outperforms individual bandit algorithms but also offers an elegant solution to practical challenges such as the 'cold start' problem. This makes our work a significant contribution to the current body of literature, showcasing an optimized, scalable, and above all, effective methodology for dynamic pricing.

TABLE IV: Margin and price selection from different CB models from ENCODE

Model	**Average Cumulative Margin**	**Average no. of times price recommendation was adopted**
LinUCB	52400	1490
CTS	52600	932
BayesUCB	51420	683
VW	50180	995
ENCODE model	61820	

A. Application of contextual bandit algorithms for dynamic pricing under different settings

In this section, we delve into our experimental observations and insights that cater to specific retail scenarios, emphasizing the challenges and innovative aspects of deploying our contextual bandit-based solutions in big data settings.

*1) **Dynamic pricing for new products**:* Pricing freshly launched products, especially in the Fashion industry, poses a unique set of challenges. Traditional methods struggle to identify suitable comparable products. To mitigate this cold-start problem, we leverage our ensemble model of contextual bandit algorithms, trained on a cluster of similar products. Our findings indicate a notable 5% margin improvement when employing the ensemble approach, as evidenced in Dataset 3 results (see Table III).

*2) **Adaptive Pricing in Competitive Landscapes**:* Our model's applicability extends to environments where competitor pricing [9] exerts a significant influence on sales and pricing strategy, as was particularly evident in Dataset 1. Uniquely, we enrich our contextual bandit algorithms with competitor-sensitive features such as effective competitor price, price differential, and leakage ratio coefficients. Leakage ratio coefficients are obtained via a two-step regression. In the first step, all important features impacting sales of a product excluding the competitor prices are regressed against sales and the residuals are obtained. In the second-step, the competitor prices are regressed against these residuals and the coefficients corresponding to the different competitor prices are called the leakage ratio coefficients. Our model dynamically adjusts sales units and margin calculations based on these metrics (refer to Equations 3,4 and 5) based on the leakage ratio coefficients. Figures 9, 10 offer insight into different competitive scenarios that our model can adeptly navigate.

What sets our approach apart is its flexibility to adapt to changing market conditions. For instance, we demonstrate the model's responsiveness to shifts in competitor pricing strategies over time, rather than adhering to a fixed strategy. A snapshot of this dynamic responsiveness is presented in Table V.

The complexity of dynamic pricing in e-commerce is multifaceted, and our novel approach leverages contextual bandit algorithms to tackle various scenarios with enhanced efficiency. Here, we elucidate on some notable settings and challenges addressed in our work.

In summary, our solution introduces a novel, data-sensitive, and adaptive approach to dynamic pricing that not only enhances margins but also provides a strategic edge in competitive markets. This is especially pertinent for new products and high-competition settings, showcasing the model's flexibility and adaptability to a wide array of retail scenarios.

*3) **Dynamic pricing driven towards Goal Seek**:* One of the primary objectives of any retailer is to achieve a desired margin goal while maximizing other business objectives in conjunction. To address this, we used our scenario modeller engine that obtains a pre-defined margin ('m') goal as an input from the retailer and employs a set of CB models to recommend optimal prices that helps the retailer attain the desired margin uplift. Refer Table VI for notations.

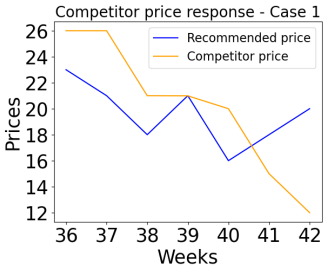

Fig. 9: Competitor response - Case 1

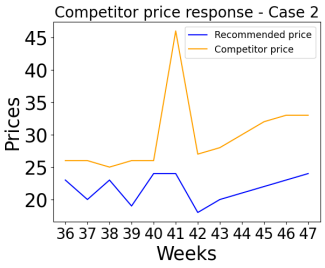

Fig. 10: Competitor response - Case 2

TABLE V: Competitor response strategy of the retailer for a particular style

Week number	Base price	Old sales	Old margin	New price	Competitor price (lagged)	New sales	New margin
1	12.79	7	37.02	12.59	11.59	9	46
2	12.59	9	45.81	12.50	10.59	9	45
3	12.50	9	45.0	12.40	10.99	10	49
4	12.40	10	49	11.59	9.5	11	45
5	11.59	10	45	11.00	9.5	12	42
6	11.00	10	42	10.79	9.5	12	39
7	10.79	11	39	10.79	10.5	16	53

To achieve this, we introduced an additional margin constraint in our dynamic pricing formulation as shown in the Equation below :

$$\frac{S_{new,i}[t] * (P_{new,i}[t] - cp_i)}{S_{new,i}[t] * P_{new,i}[t]} \geq m\% \tag{8}$$

The new objective function incorporates the above constraint as a penalty and is defined as:

$$Maximize \ (\alpha_1 * Sales + \alpha_2 * Revenue + \alpha_3 * Margin - penalty) \tag{9}$$

$$\begin{aligned} Maximize \quad & (\alpha_1 * S_{new,i}[t] + \alpha_2 \\ & * (P_{opt,i}[t] * S_{new,i}[t]) + \alpha_3 \\ & * (S_{new,i}[t] * (P_{opt,i}[t] - cp_i)) - p \\ & * m - \frac{S_{new,i}[t] * (P_{opt,i}[t] - cp_i)}{S_{new,i}[t] * P_{opt,i}[t]}) \end{aligned} \tag{10}$$

TABLE VI: Table of Notations

Notation	Description
$P_{opt,i}[t]$	Optimal price for the round 't' for style 'i'
$S_{new,i}[t]$	Sales corresponding to optimal price for round 't' for style 'i', considering inter-item & competitor effects,
cp_i	Unit cost price of style 'i'
m	Minimum margin percentage for style 'i'
p	Penalty constant for margin reduction
$\alpha_1, \alpha_2, \alpha_3$	Weights for the objective, i.e.,sales, revenue & margin such that $\alpha_1 + \alpha_2 + \alpha_3 = 1$,

The results based on this experiment are summarized in Table VII. We observe a 12.19% increase in margin based on

the new formulation set to achieve the margin goal.

TABLE VII: Goal seek scenario - results

Scenario	Average price	Total sales	Old margin	New margin	Margin change %
Without Goal seek	11.75	232661.86	753482.50	806681.48	6.59
With Goal seek	11.72	280265.89	753482.50	858144.07	12.19

*4) **Dimensionality Challenge**:* The e-commerce landscape necessitates a multitude of context features, like customer feedback and journey mapping. Our experiments revealed that the sheer number of these features imposes a computational toll. Ingeniously, we used Principal Component Analysis (PCA) to reduce this feature set, significantly decreasing algorithm training time. Refer to Tables VIII and IX for detailed comparisons.

TABLE VIII: Context features before and after the application of PCA to dataset

H&M Regular	No : of categorical features	No : of numeric features	Total No : of features	Cumulative margin
Before PCA	52	27	79	47865.67
After PCA	21	27	48	47098

TABLE IX: Running time for different datasets along with dimensionality

Dataset	Dims (d)	Samples Train	Samples Test	Rounds Train	Train time (mins)	Test time (mins)
Dataset 1	14	494	494	502	15.54	1.54
Dataset 2 before PCA	79	6154	4760	40	120	1.98
Dataset 2 after PCA	48	6154	4760	40	65.19	1.02
Dataset 3	18	413	780	515	14	0.05

*5) **Hyperparameter Tuning**:* Operationalizing bandit algorithms in real-world scenarios requires careful selection of hyperparameters [14]. Our study provides an exhaustive list of hyperparameters tailored for different data sets, contributing to computational efficiency. Specifically, we harnessed historical sales data to initialize prior means and likelihood variances, resulting in about a 20% time saving compared to random initialization.

Here, we give details about the list of hyperparameters of our bandit algorithms and details of optimal values chosen for them corresponding to different datasets. Hyperparameters 1 and 2 are very specific to the group of Bayesian algorithms as part of our CB models.

1) Prior means – The distribution of rewards corresponding to different price points (arms in our case) are initialized with reward probabilities captured from historical sales data corresponding to the different products.
2) Prior and likelihood variances – The distribution of reward uncertainty modelled via the variance of the priors and the likelihood distribution were captured from historical sales corresponding to different products.(Use of prior means, variances, and likelihood variances from historical sales $\approx$ 20% time saving – as compared to random initialization of these parameters).
3) Context dimensionality (d) – We have only considered the context features that have had higher feature importance (based on Random Forest Regressor on the context data), thus eliminating irrelevant context variables that do not significantly impact reward. ($\approx$ 50% reduction in time compared to taking all original set of features).
4) Number of arms (k) – Between the minimum and maximum bounds set for price – different granular price levels were tried. We tried price intervals [0.15, 0.25, 0.5, 0.75, 1.0, 1.5]. As we experimented with more granular price intervals, sampling from posterior distributions became computationally costly. Hence, we fixed the price intervals to 0.5.($\approx$ 2 times reduction in time after setting the right interval)
5) Time horizon (T) – The number of rounds/iterations were picked from the range [10, 40, 50, 100, 250, 500, 1000, 1500, 2000]. We introduced a stopping criterion that checks for improvements in average reward in the last 'n' rounds of training and stops training when there are no significant improvements.Without the stopping criterion, the models run upto max iterations set = 2000. ($\approx$ 4 times reduction in time, refer Table IX).

*6) **Scalability in Big Data**:* Given the vast number of products (or arms in bandit terminology), computational efficiency is crucial. We offer several algorithmic tweaks like approximate inference and caching to facilitate rapid computations, notably improving the scalability of our algorithms in big data settings. The pseudocode of our tweaked version of the Contextual Thompson Sampling algorithm is given as an example in section III.

*7) **Tweaking of existing CB algorithm to reduce the computation cost**:* In big data settings, where the number of arms (price points as per our problem setting) and data points can be very large, making our contextual bandit algorithms run efficiently is crucial. Here are some tweaks and strategies to make some of our contextual bandit algorithms run faster in such settings:

(a) Approximate Inference: One of the most time-consuming aspects of Contextual Thompson Sampling is posterior sampling for each arm's distribution. In this work, we also experimented with approximate methods like Variational Inference and Markov Chain Monte Carlo (MCMC) with fewer samples to speed up inference.
(b) Caching: Caching previously computed values, such as posterior distributions for arms, can avoid recomputing them if the data doesn't change significantly. This can

be particularly useful in situations where the data doesn't change rapidly.

(c) Model Approximation: We have used simpler models or approximations for the arm distributions, such as Gaussian approximations for modelling posteriors, priors, and likelihoods.

(d) Sampling Reduction: Instead of sampling a full distribution, we have used point estimates (i.e., mean, or median of the posterior) to approximate the arm's value in order to reduce the number of samples.

We have incorporated the above tweaks to the CB algorithms as part of our model.

*8) **Dynamic pricing for LTPs**:* LTPs [15] present a unique challenge due to their low and sparse demand. In Figure 11, it is seen that the products, once re-ordered according to the number of units sold, satisfy the well-known (long-tail) Zipf's Law [16]. Traditional pricing models fall short here. Our innovative approach integrates data analytics, customer behavior [17], and dynamic pricing to address this. Since LTPs are associated with higher production cost, we calculated revenue improvements from our model instead of margin. Metrics from dataset 1 indicate a 15.63% increase in sales volume and an 11.52% revenue increase, as summarized in Table X.

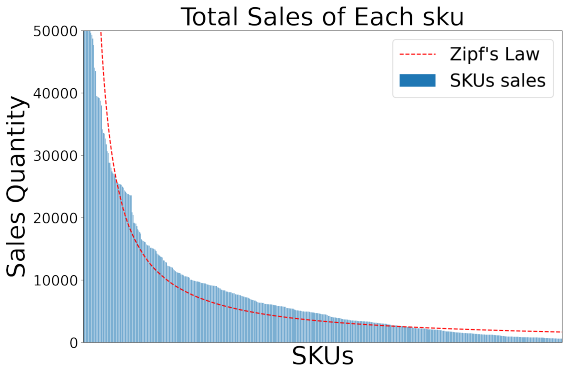

Fig. 11: Total Units sold per product - LTPs

TABLE X: LTP Pricing for Dataset 1

LTP Pricing	Sales % increase	Revenue % increase
SKU grouping	15.63	11.52
Bundling with popular SKUs	58.59	46.08
Bundling for customer segment	63.45	52.33

*9) **Dynamic pricing considering inter-item effects**:* Our algorithms also account for the complex dynamics between inter-related products, either as complements or substitutes. Incorporating these relationships as context features led to a 6% improvement in cumulative margin, as evidenced in Table XI and Figure 12.

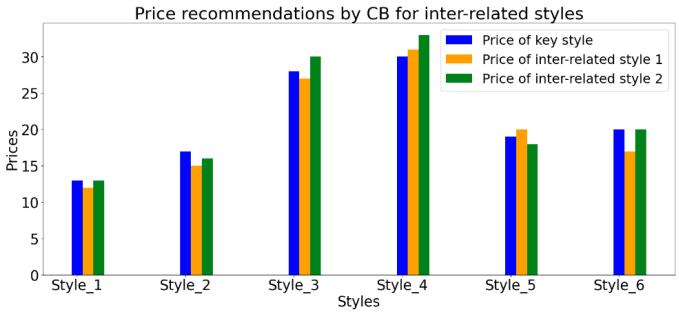

Fig. 12: Price recommendations by CB for inter-related styles

TABLE XI: Impact of inter-related styles

Dataset 1	Cumulative margin
Assuming styles are independent	12286
After consideration of inter-dependency	19651

*10) **Price family constraints**:* Ensuring that SKUs in the same price family are priced uniformly was another requirement we successfully navigated. Our context features specific to each price family helped in achieving this uniformity, as depicted in Figure 13.

Thus, our work represents a significant leap in dynamic pricing strategies by intelligently employing contextual bandit algorithms, hyperparameter tuning, and dimensionality reduction techniques to deliver robust and computationally efficient solutions. This is corroborated by the improved margins, reduced training times, and increased revenues observed across multiple datasets.

VI. Conclusion

We introduced ENCODE, a groundbreaking ensemble-based Contextual Bandit model tailored for large-scale e-commerce in the kids' clothing sector. The novelty of ENCODE lies in its integration of four diverse CB algorithms — LinUCB, Vowpal Wabbit, Contextual Thompson Sampling, and BayesUCB — to produce a comprehensive pricing strategy that optimizes both immediate and long-term margins, in accordance with retailer-defined objectives. These objectives extend until the next price adjustment, triggered by an array

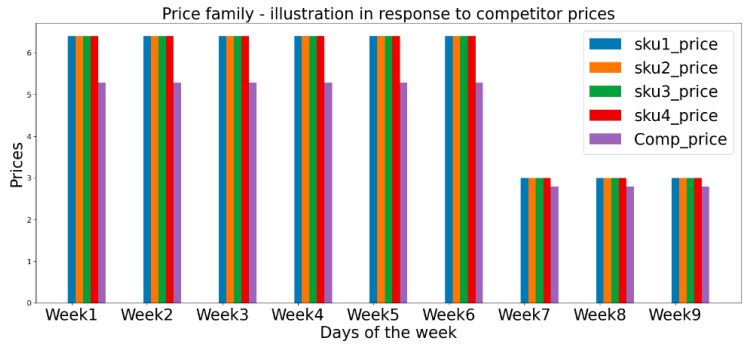

Fig. 13: Price recommendation for a price family

of variables such as competitor pricing shifts, inflation trends, and other market dynamics.

We empirically validated ENCODE's capabilities across a range of real-world scenarios, leveraging three different datasets to encompass seasonal, regular, and fashion-specific items. Notably, ENCODE accommodates dynamic variables like competitor pricing, seasonal shifts, and customer segmentation in an unprecedented manner. A focused case study revealed a 13.8% margin enhancement using Contextual Thompson Sampling alone, with an additional 6% improvement attributed to the integration of inter-related product styles. Remarkably, ENCODE surpassed the performance of individual CB models, achieving an average cumulative margin gain of 19% and a targeted margin improvement of 12.19%.

Scalability was another key focus; we successfully adapted ENCODE to manage a high volume of SKUs without compromising computational efficiency. Techniques such as dimensionality reduction, approximate inference, and caching were deployed. Utilizing advanced cloud-based computing frameworks, the model displayed robust empirical results, including a 15.63% increase in sales and an 11.52% revenue bump at the SKU group level in Dataset 1. Additionally, sales and revenues skyrocketed by 63.45% and 52.33% respectively, upon implementing bundling strategies in Dataset 1. In summary, ENCODE represents a breakthrough in the domain of dynamic pricing, offering a scalable, adaptable, deployable and empirically validated solution.

Looking ahead, we aim to delve into the realm of dynamic price personalization. Our roadmap includes broadening ENCODE's applicability across the entire product lifecycle and extending its reach to omni-channel retailing.

REFERENCES

[1] R. Kleinberg and T. Leighton, "The value of knowing a demand curve: Bounds on regret for online posted-price auctions," in *44th Annual IEEE Symposium on Foundations of Computer Science, 2003. Proceedings.* IEEE, 2003, pp. 594–605.

[2] A. V. Den Boer, "Dynamic pricing and learning: historical origins, current research, and new directions," *Surveys in operations research and management science*, vol. 20, no. 1, pp. 1–18, 2015.

[3] A. Javanmard and H. Nazerzadeh, "Dynamic pricing in high-dimensions," *The Journal of Machine Learning Research*, vol. 20, no. 1, pp. 315–363, 2019.

[4] S. Mannor and O. Shamir, "From bandits to experts: On the value of side-observations," *Advances in Neural Information Processing Systems*, vol. 24, 2011.

[5] A. Cohen, T. Hazan, and T. Koren, "Online learning with feedback graphs without the graphs," in *International Conference on Machine Learning*. PMLR, 2016, pp. 811–819.

[6] T. Lykouris, K. Sridharan, and É. Tardos, "Small-loss bounds for online learning with partial information," in *Conference on Learning Theory*. PMLR, 2018, pp. 979–986.

[7] S. Caron, B. Kveton, M. Lelarge, and S. Bhagat, "Leveraging side observations in stochastic bandits," *arXiv preprint arXiv:1210.4839*, 2012.

[8] R. Cañamares, M. Redondo, and P. Castells, "Multi-armed recommender system bandit ensembles," in *Proceedings of the 13th ACM Conference on Recommender Systems*, 2019, pp. 432–436.

[9] C. Zeng, Q. Wang, S. Mokhtari, and T. Li, "Online context-aware recommendation with time varying multi-armed bandit," in *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, 2016, pp. 2025–2034.

[10] 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.

[11] N. Gutowski, T. Amghar, O. Camp, and F. Chhel, "Gorthaur: A portfolio approach for dynamic selection of multi-armed bandit algorithms for recommendation," in *2019 IEEE 31st international conference on tools with artificial intelligence (ICTAI)*. IEEE, 2019, pp. 1164–1171.

[12] N. Patki, R. Wedge, and K. Veeramachaneni, "The synthetic data vault," in *2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*. IEEE, 2016, pp. 399–410.

[13] S. Sethuraman, U. M. G, and S. Ramanan, "Spects: Price elasticity computation using thompson sampling," in *Proceedings of the 2022 International Conference on Computational Science and Computational Intelligence (CSCI'22: December 14-16, 2022, Las Vegas, Nevada, USA)*, 2022, pp. 641–647.

[14] G. Sui and Y. Yu, "Bayesian contextual bandits for hyper parameter optimization," *IEEE Access*, vol. 8, pp. 42 971–42 979, 2020.

[15] M. Mussi, G. Genalti, F. Trovò, A. Nuara, N. Gatti, and M. Restelli, "Pricing the long tail by explainable product aggregation and monotonic bandits," in *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2022, pp. 3623–3633.

[16] R. Rousseau, "George kingsley zipf: life, ideas, his law and informetrics." *Glottometrics*, vol. 3, no. 1, pp. 11–18, 2002.

[17] S. Suresh Kumar, M. Margala, S. Siva Shankar, and P. Chakrabarti, "A novel weight-optimized lstm for dynamic pricing solutions in e-commerce platforms based on customer buying behaviour," *Soft Computing*, pp. 1–13, 2023.