# Health-Aware School Meal Recommendations with Contextual Bandits

First Author[1,2*], Second Author[2,3†] and Third Author[1,2†]

[1*]Department, Organization, Street, City, 100190, State, Country.
[2]Department, Organization, Street, City, 10587, State, Country.
[3]Department, Organization, Street, City, 610101, State, Country.

*Corresponding author(s). E-mail(s): iauthor@gmail.com;
Contributing authors: iiauthor@gmail.com; iiiauthor@gmail.com;
†These authors contributed equally to this work.

**Abstract**

This project presents the development of an open-source, data-driven recommendation and analysis tool designed to help school nutritionists, cafeteria staff, and researchers optimize meal offerings for both student preference and health outcomes. Using historical meal sales data from Fairfax County Public Schools (FCPS), the tool employs a Contextual Multi-Armed Bandit (CMAB) framework to recommend meals that balance popularity and healthiness. By incorporating contextual features such as school, time of day, and day of the week, the CMAB model—implemented through the LinUCB algorithm—learns to make personalized, context-aware meal recommendations. The reward function integrates a health weighting factor, enabling the system to account for nutritional value alongside meal popularity. The resulting methodology is embedded within an open-source library, ensuring accessibility for non-technical stakeholders. This empowers school nutrition teams to make evidence-based decisions that improve student participation and dietary quality. Beyond its immediate application, the project establishes a scalable and reproducible framework for data-driven school nutrition planning and reinforces the potential of reinforcement learning in public health and education contexts.

**Keywords:** Contextual Multi-Armed Bandit (CMAB), LinUCB, Fairfax County Public Schools (FCPS)

# 1 Introduction

School meals have a big impact on how students eat, learn, and grow. A well-balanced menu can help students stay healthy and focused, but creating one that kids actually enjoy eating isn't always easy. Popular meals are often less healthy, while nutritious options sometimes go uneaten. This makes it hard for school nutritionists and cafeteria managers to find the right balance between what students like and what's good for them.

Even though schools collect a lot of meal sales data, most cafeteria staff and nutrition professionals don't have the technical skills to analyze it or use it to guide their choices. As a result, many important insights about student food preferences and eating patterns go unused. What's needed is a simple, data-driven tool that helps non-technical users make smarter menu decisions—without requiring advanced programming or machine learning knowledge.

This project aims to build exactly that: a free and open-source meal recommendation tool powered by CMAB algorithms. Using real data from FCPS, the system learns which meals perform best under different conditions—such as the school, time of day, or day of the week. It then recommends meals that balance popularity and healthiness, helping schools improve participation while encouraging healthier eating habits.

By combining ideas from reinforcement learning and nutrition science, this tool gives school staff an easy way to make data-informed decisions. Beyond improving daily menu planning, it also creates a foundation for future research in school nutrition, making healthy choices more practical and appealing for students everywhere.

# 2 Previous Work

Prior research in CMAB has laid the foundation for adaptive decision-making in personalized and dynamic environments. Contextual bandit algorithms such as LinUCB have been extensively used across domains including recommender systems, telecommunications, and portfolio optimization. These studies demonstrate how linear contextual representations, coupled with upper confidence-based exploration strategies, enable efficient learning under uncertainty while incorporating side-information. Advances include scalable variants designed for large action spaces, methods addressing interference between units, and hybrid models integrating deep neural networks to improve reward prediction. Collectively, this body of work establishes contextual MAB as a robust framework for sequential decision-making where context features play a critical role.

In parallel, nutritional assessment research has introduced standardized nutrient-density indices such as the Nutrient-Rich Food Index 9.3 (NRF9.3). This index has been validated in multiple population-level dietary studies and is widely used to quantify overall diet quality by simultaneously rewarding beneficial nutrients and penalizing nutrients to limit. Prior work demonstrates its applicability in evaluating maternal diets, understanding associations between nutrient density and health outcomes, and comparing dietary behaviors across different populations. The NRF9.3 scoring system has also been shown to align with broader nutrient profiling methodologies, reinforcing its reliability as an evidence-based measure for capturing the quality of food choices. These contributions establish NRF9.3 as a scientifically grounded index for assessing nutrient adequacy in diverse nutritional contexts.

## 3 Methodology

The proposed system follows a unified data-to-recommendation pipeline designed to transform raw school meal records into adaptive, health-aware menu suggestions. The process begins with ingesting the full historical dataset, which includes meal identifiers, nutrient attributes, serving dates, school metadata, and student selection frequencies. These records are cleaned, merged, and transformed into a contextual feature matrix capturing nutritional density, popularity indicators, categorical encodings, and temporal attributes such as weekday and seasonal trends. To quantify the health dimension of meals, each item is scored using the NRF9.3, producing a continuous metric that reflects overall nutrient density relative to recommended intake thresholds. Popularity is computed from empirical selection frequencies and normalized to serve as a second reward component. Both health and popularity signals are then combined into a weighted reward formulation that allows tuning the relative importance of nutritional quality versus student preference.

After preprocessing, all features and rewards are scaled and aligned into the contextual bandit framework. Each meal item constitutes an arm, and the context vector for each decision point incorporates the full set of nutritional, temporal, and categorical attributes. The system employs a LinUCB algorithm, which models rewards as a linear function of these contextual features while maintaining uncertainty estimates for principled exploration. At each time step, LinUCB selects the meal with the highest upper confidence bound, balancing exploitation of historically strong items with exploration of potentially underutilized but nutritionally beneficial options. The bandit is trained and evaluated by simulating sequential decisions over the historical timeline, updating model parameters after each observed reward. To make the pipeline accessible for deployment, all data processing, model training, and evaluation steps are fully automated in Python, with outputs written to reproducible CSVs and visualized through exploratory analyses.

To contextualize the performance of the contextual bandit approach, two baseline recommenders are implemented for comparison: a rule-based recommender and a random recommender. The rule-based system selects the meal with the highest health score (NRF9.3) at each decision point, independent of contextual or historical reward signals, effectively modeling a purely health-driven policy. In contrast, the random recommender selects meals uniformly at random, providing a lower-bound benchmark for non-informed decision-making. Both baselines are evaluated under the same sequential simulation environment as the bandit model, allowing direct computation of cumulative rewards and regret relative to the optimal counterfactual reward trajectory. These baselines serve to illustrate the advantage of contextual decision-making, demonstrating how the bandit approach achieves substantially lower regret by simultaneously leveraging nutritional information, contextual patterns, and adaptive exploration.

# 4 Dataset

## 4.1 Cafeteria Sales Dataset

The primary dataset used in this study consists of detailed cafeteria sales transactions obtained from FCPS. Each row represents a specific meal item sold at a given school, on a particular date, and during a defined meal period (e.g., breakfast or lunch). The dataset includes contextual attributes such as `time_of_day`, `school_code`, `school_name`, and `date`, which enable both temporal and school-level analyses.

For each item, the dataset reports the total quantity sold as well as a comprehensive breakdown across student meal programs, including free, reduced-price, and full-price meals. Additional fields capture adult purchases, à-la-carte transactions (`alac_student`, `alac_adult`), and earned revenue categories (`earned_student`, `earned_adult`, `earned_alac_student`, `earned_alac_adult`). Administrative adjustments are recorded under `adj_alac` and `adj_meal`.

This granular structure allows the dataset to reflect both reimbursable meals and à-la-carte demand, supporting precise feature construction for the contextual multi-armed bandit model. The dataset spans multiple days and schools, offering sufficient variability for robust training and evaluation.

**Table 1** Description of Columns in the FCPS Cafeteria Sales Dataset

| **Column Name** | **Description** |
|---|---|
| `time_of_day` | Meal period during which the item was served (e.g., breakfast, lunch). |
| `school_code` | Numeric identifier assigned to each school by FCPS. |
| `school_name` | Full name of the school where the transaction occurred. |
| `date` | Calendar date of the meal service (MM/DD/YYYY). |
| `item` | Internal item code representing a specific meal or à-la-carte product. |
| `description` | Human-readable label describing the food item. |
| `total` | Total units of the item sold on that date and meal period. |
| `free_meals` | Reimbursable meals provided to students eligible for free meals. |
| `reduced_price_meals` | Reimbursable meals provided to students eligible for reduced-price meals. |
| `full_price_meals` | Reimbursable meals purchased at full price. |
| `adults` | Meals or items purchased by adults. |
| `alac_student` | À-la-carte items purchased by students. |
| `alac_adult` | À-la-carte items purchased by adults. |
| `earned_student` | Revenue-earning meals attributed to student purchases. |
| `earned_adult` | Revenue-earning meals attributed to adult purchases. |
| `earned_alac_student` | Revenue-earning à-la-carte items purchased by students. |
| `earned_alac_adult` | Revenue-earning à-la-carte items purchased by adults. |
| `adj_alac` | Administrative adjustments applied to à-la-carte counts. |
| `adj_meal` | Administrative adjustments applied to meal counts. |

## 4.2 Nutrition Enriched Item Dataset

The second dataset integrates meal sales records with detailed nutritional information for each food item. Each row corresponds to a unique menu item (e.g., Cereal Meal, Mini Pancakes, PBJ Power Pack) and contains nutrient-level attributes derived from linqconnect API . These features include energy content, macronutrients (protein, carbohydrates, sugars, fats), micronutrients (calcium, iron, sodium), and serving-size metadata. The dataset also includes derived variables such as diet and religious restrictions, number of candidate items mapped to each sales entry, and indicator flags specifying whether full nutrient information was available. This dataset enables the construction of a comprehensive feature matrix for contextual bandit modeling, allowing the system to learn relationships between sales behavior and nutritional quality. The combination of both sales and nutrient information is essential for designing a health-aware reward function that balances student preference and meal healthfulness.

**Table 2** Column Descriptions for the Sales–Nutrition Merged Dataset

| **Column** | **Description** |
|---|---|
| `sales_item` | Original meal item name from FCPS sales data. |
| `mapped_items` | List of matched USDA or manufacturer nutrition items used to derive nutrient values. |
| `GramsPerServing` | Total grams per serving of the matched item. |
| `Calories` | Energy content per serving (kcal). |
| `Protein` | Protein content (g). |
| `Total_Carbohydrate` | Total carbohydrates per serving (g). |
| `Dietary_Fiber` | Fiber content (g). |
| `Total_Sugars` | Total sugars per serving (g). |
| `Added_Sugars` | Added sugars per serving (g). |
| `Total_Fat` | Total fat content (g). |
| `Saturated_Fat` | Saturated fat content (g). |
| `Trans_Fat` | Trans fat content (g). |
| `Cholesterol` | Cholesterol content (mg). |
| `Sodium` | Sodium content (mg). |
| `Calcium` | Calcium content (mg). |
| `Iron` | Iron content (mg). |
| `Potassium` | Potassium content (mg). |
| `VitaminD` | Vitamin D content (µg). |
| `HasNutrients` | Indicator variable (1/0) denoting whether full nutrient information was available. |
| `DietaryRestrictions` | Dietary restriction tags (e.g., gluten-free, vegetarian), if applicable. |
| `ReligiousRestrictions` | Religious dietary restriction tags, if applicable. |
| `method` | Aggregation method used to compute nutrient values across matched candidates (e.g., `median_over_topK`). |
| `n_candidates` | Number of matched candidate nutrition items used for aggregation. |
| `sales_volume` | Total sales volume associated with this item across the full dataset. |

## 4.3 Health Score Computation

To evaluate the nutritional quality of each meal item, we employ the Nutrient Rich Foods Index (NRF9.3), a standardized nutrient-density metric. The NRF9.3 score balances nutrients to encourage against nutrients to limit. Beneficial nutrients include protein, dietary fiber, vitamin D, calcium, iron, potassium, vitamin A, and vitamin C, while limiting nutrients consist of added sugars, saturated fat, and sodium.

For each nutrient, we compute the percentage of the USDA Daily Value (DV) contained in a single serving. The NRF9.3 score for an item is then calculated as:

$$\mathrm{NRF9.3} = \sum_{\text{good}} \%\mathrm{DV} \; - \; \sum_{\text{bad}} \%\mathrm{DV}.$$

Because raw NRF9.3 values may vary widely across food categories and school levels, we normalize the scores to a fixed range of 0–10 to ensure comparability. Daily Values are adjusted according to the nutritional requirements of elementary, middle, and high school students based on USDA guidelines. This standardized health score provides a consistent and interpretable measure of the nutritional quality of meals served across the district.

## 4.4 Fully Merged Sales–Nutrition–Health Dataset

The third dataset is a fully merged, transaction-level table that combines the original FCPS sales records with item-level nutritional information and derived health scores. Each row corresponds to a specific food item sold at a particular school, on a given date, and during a given meal period (e.g., breakfast). In addition to the operational sales fields (time of day, school identifiers, item codes, quantities sold, and program breakdowns such as free, reduced-price, and full-price meals), the dataset includes the mapped nutrient profile for each item, school-level grouping (e.g., elementary), and a continuous health score.

Overall, the merged dataset contains 224536 rows and 47 columns. The rows represent individual school–date–item combinations, while the 47 columns span four categories: (i) sales and program participation fields, (ii) item identifiers and mapped nutrition references, (iii) detailed nutrient values (energy, macronutrients, micronutrients, and related dietary indicators), and (iv) derived attributes such as health score, school group, and aggregation metadata. This rich structure allows the learning algorithm to jointly model demand, nutritional quality, and school context at the transaction level.

**Table 3**: Column Descriptions for the Fully Merged Dataset

| **Column** | **Description** |
|---|---|
| `time_of_day` | Meal session (breakfast or lunch). |
| `school_code` | Numerical identifier for each school. |
| `school_name` | Full name of the school where the sale occurred. |
| `date` | Date of transaction (MM/DD/YYYY). |

| Column | Description |
|---|---|
| item | Internal FCPS item code. |
| description | Human-readable name of the food item. |
| total | Total quantity of the item sold. |
| free_meals | Count of free reimbursable meals. |
| reduced_price_meals | Count of reduced-price reimbursable meals. |
| full_price_meals | Count of full-price reimbursable meals. |
| adults | Adult purchases of the item. |
| alac_student | À-la-carte purchases by students. |
| alac_adult | À-la-carte purchases by adults. |
| earned_student | Earned revenue from student meals. |
| earned_adult | Earned revenue from adult meals. |
| earned_alac_student | Earned revenue from student à-la-carte purchases. |
| earned_alac_adult | Earned revenue from adult à-la-carte purchases. |
| adj_alac | Administrative adjustment to à-la-carte counts. |
| adj_meal | Administrative adjustment to meal counts. |
| sales_item | Original item label used in sales data. |
| mapped_items | USDA/manufacturer nutrient items mapped to this sales item. |
| GramsPerServing | Serving size in grams. |
| Calories | Energy per serving (kcal). |
| Protein | Protein content (g). |
| Total Carbohydrate | Total carbohydrates (g). |
| Dietary Fiber | Dietary fiber (g). |
| Total Sugars | Total sugars (g). |
| Added Sugars | Added sugars (g). |
| Total Fat | Total fat content (g). |
| Saturated Fat | Saturated fat content (g). |
| Trans Fat | Trans-fat content (g). |
| Cholesterol | Cholesterol (mg). |
| Sodium | Sodium (mg). |
| Vitamin D (D2 + D3) | Vitamin D content (µg). |
| Calcium | Calcium (mg). |
| Iron | Iron (mg). |
| Potassium | Potassium (mg). |
| Vitamin A | Vitamin A content (IU or µg). |
| Vitamin C | Vitamin C content (mg). |
| HasNutrients | Indicator for nutrient completeness (1 = has full nutrients). |
| DietaryRestrictions | Dietary restriction tags (e.g., gluten-free, vegetarian). |
| ReligiousRestrictions | Religious dietary restriction tags. |
| method | Aggregation method used (e.g., median_over_topK). |
| n_candidates | Number of nutrition matches used for aggregation. |
| sales_volume | Total sales volume across the dataset. |
| school_group | School classification (elementary, middle, high). |

| Column | Description |
|---|---|
| `HealthScore` | Computed health score used in reward formulation. |

## 4.5 Bandit Framework

We formulate the school meal recommendation problem as a Contextual Multi-Armed Bandit (CMAB) task, which enables sequential decision-making under uncertainty. Unlike traditional supervised learning models that rely solely on historical observations, the CMAB framework supports interactive learning: at each time step, the agent observes a context, selects an action, receives a reward, and updates its policy based on the observed feedback.

In this setting, the *context* consists of school- and time-specific attributes, including the school code, meal period, date, and day of the week. The *action space* corresponds to the set of meal items available during a given time slot. To jointly capture student preference and nutritional value, the *reward* is defined as:

$$r_t = \text{total meals served} + \lambda \cdot \text{health score}_t$$

where $\lambda$ controls the trade-off between popularity and nutritional quality.

This formulation allows the agent to balance *exploitation*—recommending meals known to perform well—and *exploration*—testing alternative items that may yield improved outcomes. By incorporating the NRF9.3-based health score directly into the reward signal, the CMAB approach supports adaptive, data-driven menu recommendations that aim to improve both student participation and meal healthfulness across diverse schools.

## 4.6 Action Space Definition

In the contextual bandit setting, the action space corresponds to the set of meal items that may be recommended at a given decision point. The FCPS dataset contains more than 160 unique items, including entrées, sides, beverages, and à-la-carte options. Because menu availability varies across schools, dates, and meal periods, not all items are feasible at every time step.

To capture these operational constraints, we construct an *action matrix* that records, for each time slot, the set of items that were available for service. This dynamic action space ensures that the bandit agent selects only actions that reflect actual menu offerings and aligns the decision process with real cafeteria conditions.

## 4.7 Context Construction

The contextual features describe the environment in which each recommendation is made and include both school-level and temporal attributes. Specifically, the context vector incorporates the school code, meal period (breakfast or lunch), date, and the corresponding day of the week. These variables capture systematic differences in meal preference across schools as well as temporal fluctuations driven by schedule patterns and seasonal trends.

By providing the agent with a rich representation of the decision environment, the contextual structure enables the model to learn how meal uptake varies across settings and to generate recommendations tailored to specific schools and time periods.

## 4.8 Reward Function

To balance student preference with nutritional quality, we design a reward function that integrates both sales performance and the health score of each meal. The reward for item $a$ at time $t$ is defined as:

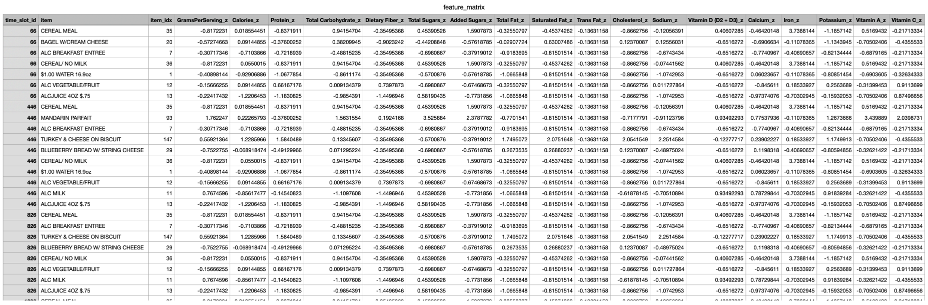
$$r_t = \text{sales}_t + \lambda \cdot \text{health score}_t$$

where the total number of units sold reflects student uptake and the health score is derived from the NRF9.3 index. The parameter $\lambda$ modulates the trade-off between popularity and nutrition: higher values emphasize healthier meals, while $\lambda = 0$ yields a purely preference-based strategy.

This reward structure enables the bandit agent to learn recommendations that promote both meal appeal and nutritional value, aligning with the broader objectives of FCPS meal planning.

## 4.9 Feature Matrix Overview

Figure 1 shows the structure of the feature matrix used in the LinUCB model. Each row represents a specific (item, context) instance, while the columns encode nutritional variables and contextual attributes. This combined representation enables the model to learn context-dependent reward predictions.



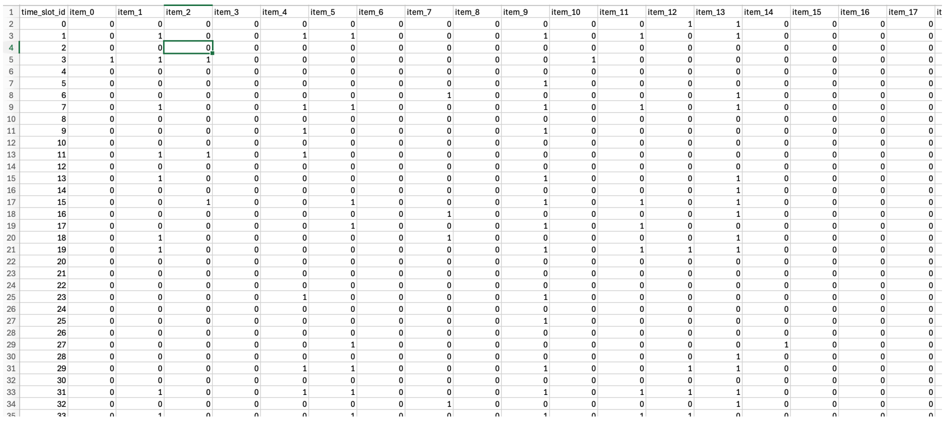
feature_matrix

| time_slot_id | item | item_idx | GramsPerServing_z | Calories_z | Protein_z | Total Carbohydrate_z | Dietary Fiber_z | Total Sugars_z | Added Sugars_z | Total Fat_z | Saturated Fat_z | Trans Fat_z | Cholesterol_z | Sodium_z | Vitamin D (D2 + D3)_z | Calcium_z | Iron_z | Potassium_z | Vitamin A_z | Vitamin C_z |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 66 | CEREAL MEAL | 35 | -0.8172231 | 0.018554451 | -0.8371911 | 0.94154704 | -0.35495368 | 0.45390528 | 1.5907873 | -0.32550797 | -0.45374262 | -0.13631158 | -0.8662756 | -0.12056391 | 0.40607285 | -0.46420148 | 3.7388144 | -1.1857142 | 0.5169432 | -0.21713334 |
| 66 | BAGEL W/CREAM CHEESE | 20 | -0.57274663 | 0.09144855 | -0.37600252 | 0.38209945 | -0.9023242 | -0.44208848 | -0.57618785 | -0.02907724 | 0.63007486 | -0.13631158 | 0.12370087 | 0.12556031 | -0.6516272 | -0.6906634 | -0.11078365 | -1.1343945 | -0.70502406 | -0.4355533 |
| 66 | ALC BREAKFAST ENTREE | 7 | -0.30717346 | -0.7103866 | -0.7218939 | -0.48815235 | -0.35495368 | -0.6980867 | -0.37919012 | -0.9183695 | -0.81501514 | -0.13631158 | -0.8662756 | -0.6743434 | -0.6516272 | -0.7740967 | -0.40690657 | -0.82134444 | -0.6879165 | -0.21713334 |
| 66 | CEREAL/ NO MILK | 36 | -0.8172231 | 0.0550015 | -0.8371911 | 0.94154704 | -0.35495368 | 0.45390528 | 1.5907873 | -0.32550797 | -0.45374262 | -0.13631158 | -0.8662756 | -0.07441562 | 0.40607285 | -0.46420148 | 3.7388144 | -1.1857142 | 0.5169432 | -0.21713334 |
| 66 | $1.00 WATER 16.9oz | 1 | -0.40898144 | -0.92906886 | -1.0677854 | -0.8611174 | -0.35495368 | -0.5700876 | -0.57618785 | -1.0665848 | -0.81501514 | -0.13631158 | -0.8662756 | -1.0742953 | -0.6516272 | 0.06023657 | -0.11078365 | -0.80851454 | -0.6903605 | -0.32634333 |
| 66 | ALC VEGETABLE/FRUIT | 12 | -0.15666255 | 0.09144855 | 0.66167176 | 0.009134379 | 0.7397873 | -0.6980867 | -0.67468673 | -0.32550797 | -0.81501514 | -0.13631158 | -0.8662756 | 0.011727864 | -0.6516272 | -0.845611 | 0.18533927 | 0.2563689 | -0.31399453 | 0.9113699 |
| 66 | ALCJUICE 4OZ $.75 | 13 | -0.22417432 | -1.2206453 | -1.1830825 | -0.9854391 | -1.4496946 | 0.58190435 | -0.7731856 | -1.0665848 | -0.81501514 | -0.13631158 | -0.8662756 | -1.0742953 | -0.6516272 | -0.97374076 | -0.70302945 | -0.15932053 | -0.70502406 | 0.87496656 |
| 446 | CEREAL MEAL | 35 | -0.8172231 | 0.018554451 | -0.8371911 | 0.94154704 | -0.35495368 | 0.45390528 | 1.5907873 | -0.32550797 | -0.45374262 | -0.13631158 | -0.8662756 | -0.12056391 | 0.40607285 | -0.46420148 | 3.7388144 | -1.1857142 | 0.5169432 | -0.21713334 |
| 446 | MANDARIN PARFAIT | 93 | 1.762247 | 0.22265793 | -0.37600252 | 1.5631554 | 0.1924168 | 3.525884 | 2.3787782 | -0.7701541 | -0.81501514 | -0.13631158 | -0.7177791 | -0.91123796 | 0.93492293 | 0.77537936 | -0.11078365 | 1.2673666 | 3.439889 | 2.0398731 |
| 446 | ALC BREAKFAST ENTREE | 7 | -0.30717346 | -0.7103866 | -0.7218939 | -0.48815235 | -0.35495368 | -0.6980867 | -0.37919012 | -0.9183695 | -0.81501514 | -0.13631158 | -0.8662756 | -0.6743434 | -0.6516272 | -0.7740967 | -0.40690657 | -0.82134444 | -0.6879165 | -0.21713334 |
| 446 | TURKEY & CHEESE ON BISCUIT | 147 | 0.55921364 | 1.2285966 | 1.5840489 | 0.13345607 | -0.35495368 | -0.5700876 | -0.37919012 | 1.7495072 | 2.0751648 | -0.13631158 | 2.0541549 | 2.2514584 | -0.12277717 | 0.23902227 | 0.18533927 | 1.1749913 | -0.70502406 | -0.4355533 |
| 446 | BLUEBERRY BREAD W/ STRING CHEESE | 29 | -0.7522755 | -0.068918474 | -0.49129966 | 0.071295224 | -0.35495368 | -0.6980867 | -0.57618785 | 0.2673535 | 0.26880237 | -0.13631158 | 0.12370087 | -0.48975024 | -0.6516272 | 0.1198318 | -0.40690657 | -0.80594856 | -0.32621422 | -0.21713334 |
| 446 | CEREAL/ NO MILK | 36 | -0.8172231 | 0.0550015 | -0.8371911 | 0.94154704 | -0.35495368 | 0.45390528 | 1.5907873 | -0.32550797 | -0.45374262 | -0.13631158 | -0.8662756 | -0.07441562 | 0.40607285 | -0.46420148 | 3.7388144 | -1.1857142 | 0.5169432 | -0.21713334 |
| 446 | $1.00 WATER 16.9oz | 1 | -0.40898144 | -0.92906886 | -1.0677854 | -0.8611174 | -0.35495368 | -0.5700876 | -0.57618785 | -1.0665848 | -0.81501514 | -0.13631158 | -0.8662756 | -1.0742953 | -0.6516272 | 0.06023657 | -0.11078365 | -0.80851454 | -0.6903605 | -0.32634333 |
| 446 | ALC VEGETABLE/FRUIT | 12 | -0.15666255 | 0.09144855 | 0.66167176 | 0.009134379 | 0.7397873 | -0.6980867 | -0.67468673 | -0.32550797 | -0.81501514 | -0.13631158 | -0.8662756 | 0.011727864 | -0.6516272 | -0.845611 | 0.18533927 | 0.2563689 | -0.31399453 | 0.9113699 |
| 446 | ALC MILK | 11 | 0.7674596 | -0.85617477 | -0.14540823 | -1.1097608 | -1.4496946 | 0.45390528 | -0.7731856 | -1.0665848 | -0.81501514 | -0.13631158 | -0.61878145 | -0.70510894 | 0.93492293 | 0.78729844 | -0.70302945 | 0.91839284 | -0.32621422 | -0.4355533 |
| 446 | ALCJUICE 4OZ $.75 | 13 | -0.22417432 | -1.2206453 | -1.1830825 | -0.9854391 | -1.4496946 | 0.58190435 | -0.7731856 | -1.0665848 | -0.81501514 | -0.13631158 | -0.8662756 | -1.0742953 | -0.6516272 | -0.97374076 | -0.70302945 | -0.15932053 | -0.70502406 | 0.87496656 |
| 826 | CEREAL MEAL | 35 | -0.8172231 | 0.018554451 | -0.8371911 | 0.94154704 | -0.35495368 | 0.45390528 | 1.5907873 | -0.32550797 | -0.45374262 | -0.13631158 | -0.8662756 | -0.12056391 | 0.40607285 | -0.46420148 | 3.7388144 | -1.1857142 | 0.5169432 | -0.21713334 |
| 826 | ALC BREAKFAST ENTREE | 7 | -0.30717346 | -0.7103866 | -0.7218939 | -0.48815235 | -0.35495368 | -0.6980867 | -0.37919012 | -0.9183695 | -0.81501514 | -0.13631158 | -0.8662756 | -0.6743434 | -0.6516272 | -0.7740967 | -0.40690657 | -0.82134444 | -0.6879165 | -0.21713334 |
| 826 | TURKEY & CHEESE ON BISCUIT | 147 | 0.55921364 | 1.2285966 | 1.5840489 | 0.13345607 | -0.35495368 | -0.5700876 | -0.37919012 | 1.7495072 | 2.0751648 | -0.13631158 | 2.0541549 | 2.2514584 | -0.12277717 | 0.23902227 | 0.18533927 | 1.1749913 | -0.70502406 | -0.4355533 |
| 826 | BLUEBERRY BREAD W/ STRING CHEESE | 29 | -0.7522755 | -0.068918474 | -0.49129966 | 0.071295224 | -0.35495368 | -0.6980867 | -0.57618785 | 0.2673535 | 0.26880237 | -0.13631158 | 0.12370087 | -0.48975024 | -0.6516272 | 0.1198318 | -0.40690657 | -0.80594856 | -0.32621422 | -0.21713334 |
| 826 | CEREAL/ NO MILK | 36 | -0.8172231 | 0.0550015 | -0.8371911 | 0.94154704 | -0.35495368 | 0.45390528 | 1.5907873 | -0.32550797 | -0.45374262 | -0.13631158 | -0.8662756 | -0.07441562 | 0.40607285 | -0.46420148 | 3.7388144 | -1.1857142 | 0.5169432 | -0.21713334 |
| 826 | ALC VEGETABLE/FRUIT | 12 | -0.15666255 | 0.09144855 | 0.66167176 | 0.009134379 | 0.7397873 | -0.6980867 | -0.67468673 | -0.32550797 | -0.81501514 | -0.13631158 | -0.8662756 | 0.011727864 | -0.6516272 | -0.845611 | 0.18533927 | 0.2563689 | -0.31399453 | 0.9113699 |
| 826 | ALC MILK | 11 | 0.7674596 | -0.85617477 | -0.14540823 | -1.1097608 | -1.4496946 | 0.45390528 | -0.7731856 | -1.0665848 | -0.81501514 | -0.13631158 | -0.61878145 | -0.70510894 | 0.93492293 | 0.78729844 | -0.70302945 | 0.91839284 | -0.32621422 | -0.4355533 |
| 826 | ALCJUICE 4OZ $.75 | 13 | -0.22417432 | -1.2206453 | -1.1830825 | -0.9854391 | -1.4496946 | 0.58190435 | -0.7731856 | -1.0665848 | -0.81501514 | -0.13631158 | -0.8662756 | -1.0742953 | -0.6516272 | -0.97374076 | -0.70302945 | -0.15932053 | -0.70502406 | 0.87496656 |

**Fig. 1** Snapshot of the constructed feature matrix integrating nutritional and contextual attributes.

## 4.10 Action Matrix Overview

The action matrix encodes the set of menu items available during each time slot. Each row corresponds to a unique decision point defined by the school, date, and meal period, while each column represents a specific meal item. A binary indicator specifies whether the item was served at that time. This structure ensures that the bandit agent selects only feasible actions that reflect actual cafeteria offerings.



| 1 | time_slot_id | item_0 | item_1 | item_2 | item_3 | item_4 | item_5 | item_6 | item_7 | item_8 | item_9 | item_10 | item_11 | item_12 | item_13 | item_14 | item_15 | item_16 | item_17 | it |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | |
| 3 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | |
| 4 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 5 | 3 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 6 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 7 | 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 8 | 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | |
| 9 | 7 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | |
| 10 | 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 11 | 9 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 12 | 10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 13 | 11 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 14 | 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 15 | 13 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | |
| 16 | 14 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | |
| 17 | 15 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | |
| 18 | 16 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | |
| 19 | 17 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 20 | 18 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | |
| 21 | 19 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | |
| 22 | 20 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 23 | 21 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 24 | 22 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 25 | 23 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 26 | 24 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 27 | 25 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 28 | 26 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 29 | 27 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | |
| 30 | 28 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | |
| 31 | 29 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | |
| 32 | 30 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 33 | 31 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | |
| 34 | 32 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | |
| 35 | 33 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | |

**Fig. 2** Snapshot of the action matrix showing availability of meal items across time slots.

# 5 Models

## 5.1 Why LinUCB?

Choosing an appropriate learning framework is critical for developing a recommendation system that balances student preferences with nutritional goals. Traditional supervised learning approaches are limited in this setting because they rely solely on historical data and lack the ability to make sequential decisions or balance exploration and exploitation. They cannot adapt to evolving student behavior or account for uncertainty in rarely served items. Deep reinforcement learning methods, while expressive, require substantial interaction data, demand significant computational resources, and offer limited interpretability—an important consideration in school nutrition applications.

The LinUCB algorithm provides a favorable compromise between efficiency, interpretability, and learning effectiveness. As a contextual bandit method, LinUCB incorporates school-level and temporal attributes, along with nutritional features, to generate context-dependent recommendations. Its upper-confidence-bound formulation enables systematic exploration of underutilized but potentially healthy meal options while exploiting items known to perform well.

LinUCB is also computationally efficient, scaling to more than 160 items and tens of thousands of time slots in the FCPS dataset. Its linear model structure produces transparent parameter estimates that are easily interpretable by administrators and nutrition specialists. These characteristics make LinUCB a practical and principled choice for real-time, health-aware meal recommendation in public school settings.

## 5.2 LinUCB Algorithm

The LinUCB algorithm provides an efficient solution to the contextual multi-armed bandit problem by assuming that rewards can be expressed as a linear function of item features. For each arm $a$, the algorithm maintains a covariance matrix $A_a$ and a response vector $b_a$, both of which are updated incrementally as new observations are collected.

The parameter vector for arm $a$ is estimated as

$$\theta_a = A_a^{-1} b_a,$$

representing the learned linear mapping between features and expected rewards. Given a context vector $x_{t,a} \in \mathbb{R}^d$, the algorithm computes an upper-confidence-bound score:

$$p_{t,a} = \theta_a^\top x_{t,a} + \alpha \sqrt{x_{t,a}^\top A_a^{-1} x_{t,a}},$$

where the first term encourages exploitation of high-performing items and the second term promotes exploration based on model uncertainty. The hyperparameter $\alpha$ controls the trade-off between these behaviors.

After selecting the arm with the highest score, observing the reward $r_t$, and receiving the corresponding feature vector, the model updates its statistics:

$$A_a \leftarrow A_a + x_{t,a} x_{t,a}^\top, \qquad b_a \leftarrow b_a + r_t x_{t,a}.$$

The LinUCB framework offers advantages well suited to the FCPS meal recommendation task. Its linear structure is computationally efficient, its exploration mechanism enables evaluation of lesser-known but potentially healthier options, and its arm-specific representation preserves interpretability while incorporating context-dependent reward prediction.

## 5.3 Model Training Procedure

The training process for the LinUCB agent follows a structured pipeline that converts the FCPS data into contextual bandit tensors and iteratively updates the model based on observed rewards. Training begins with the computation of rewards, where both sales and health scores are scaled to the range $[0, 10]$ and combined according to

$$r = \text{sales}_{\text{scaled}} + \lambda \cdot \text{health}_{\text{scaled}}.$$

This formulation ensures that popularity and nutrition contribute comparably to the reward signal.

To support contextual bandit learning, the feature matrix, action matrix, and metadata are transformed into three-dimensional tensors using the `_build_bandit_tensors` routine. For each time slot, this routine produces:

1. a feature tensor $data \in \mathbb{R}^{T \times A \times d}$ describing item features,
2. a reward tensor $rewards \in \mathbb{R}^{T \times A}$ containing realized rewards, and
3. an availability mask $mask \in \{0, 1\}^{T \times A}$ indicating which items were offered.

This representation allows the agent to learn under realistic menu constraints.

During training, the LinUCB model iterates through time steps $t = 1, \ldots, T$. At each step, the agent observes the available arms, computes an upper-confidence-bound score for each feasible item, selects the arm with the highest score, and updates the corresponding parameters:

$$A_a \leftarrow A_a + x_{t,a} x_{t,a}^\top, \qquad b_a \leftarrow b_a + r_t x_{t,a}.$$

The algorithm tracks cumulative reward, oracle reward, regret, and average reward to quantify learning performance.

After training, all learned parameters—including the arm-specific matrices $A_a$ and $b_a$, exploration statistics, and model metadata—are saved using the `save_model` function. The complete lifecycle therefore consists of (i) reward construction, (ii) tensor building, (iii) iterative bandit training, and (iv) model persistence, enabling reproducible and deployable meal recommendation models across FCPS.

## 5.4 System Architecture Overview

The proposed health-aware meal recommendation system integrates contextual data, nutritional information, and bandit learning into a unified architecture. As illustrated in Figure 3, the pipeline begins with the extraction of contextual variables such as school code, date, and meal period. These variables characterize the decision environment and define the conditions under which recommendations are generated.

The environment module constructs two core components: (i) a feature matrix encoding nutritional and contextual attributes for each meal item, and (ii) an action matrix specifying which items were available during each time slot. These components define the feasible action space and ensure that recommendations adhere to real-world FCPS menu constraints.

At each step, the LinUCB agent computes an upper-confidence-bound score for all available items and selects the arm with the highest score. After observing the realized reward—calculated from scaled sales and health scores—the agent updates its internal parameters, refining both its reward estimates and uncertainty terms.

This structure establishes a closed feedback loop in which context shapes the environment, the agent selects an action, the environment returns a reward, and the agent updates its model accordingly. The architecture enables adaptive, data-driven recommendations that balance student preferences with nutritional goals.

## 5.5 Random Baseline Model

To benchmark the performance of the proposed LinUCB agent, we include a non-learning random policy as a baseline. The random model does not estimate expected rewards or exploit structure in the feature space; instead, it selects an available meal item uniformly at random at each decision point. This yields a naive strategy that is agnostic to both historical outcomes and nutritional information and thus serves as a lower bound for more sophisticated methods.

The random policy operates within the same environment and uses the same reward formulation as the other models. For each time step $t$, the set of feasible actions is determined by the availability mask derived from the action matrix, which encodes which items were actually served during that time slot. The agent samples an arm $a$

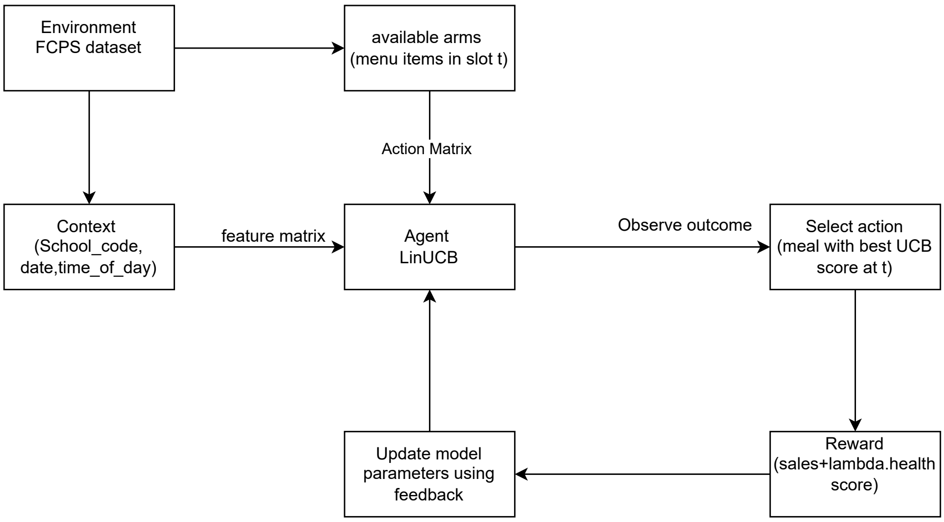


**Fig. 3** System architecture for the contextual bandit–based meal recommendation framework. The flow diagram is adapted from the project design.

uniformly from this set and receives a reward

$$r_t = \text{sales}_{t,\text{scaled}} + \lambda \cdot \text{health}_{t,\text{scaled}},$$

where both sales and health scores are scaled to the range $[0, 10]$. The random agent does not maintain or update any parameters across time, so its behavior remains stationary throughout the experiment.

Including this baseline serves two purposes. First, it quantifies the advantage of leveraging contextual and nutritional features by comparing informed policies ( LinUCB) against uninformed selection. Second, it provides a reference point for cumulative reward and regret: any deployable policy should substantially outperform the random model.

## 5.6 Health-First Baseline Model

The health-first model provides a baseline policy that prioritizes nutritional quality above all other considerations. Unlike learning-based approaches, this policy does not estimate reward functions or update parameters over time. Instead, it applies a deterministic rule in which, for each decision point, the item(s) with the highest nutritional health score are selected.

To implement this rule, all rows in the dataset are grouped by their corresponding time slot, ensuring that each decision is evaluated independently. Within each group, the available items are ranked in descending order by health score. In cases where multiple items share the same health score, we apply a secondary tie-breaking criterion using the $\lambda$-reward (a combination of scaled sales and scaled health). This avoids row-order bias and ensures that the selection remains stable and aligned with the evaluation metric used for all models.

After selecting the top-$K$ healthiest items for each time slot, we compute cumulative reward using the same $\lambda$-reward employed by the Random, and LinUCB models. For comparability, the oracle reward for each time slot is defined as the sum of the top-$K$ items under the $\lambda$-reward, and regret is calculated accordingly. Because the health-first model does not make use of sales patterns, feature representations, or exploration–exploitation dynamics, its performance highlights the trade-off between maximizing nutritional quality and optimizing overall reward. In our experiments, the health-first rule achieves strong nutritional outcomes but exhibits higher cumulative regret compared to LinUCB, illustrating the importance of balancing health and popularity rather than optimizing either dimension in isolation.

## 5.7 Evaluation Metrics

We assess model performance using a set of metrics that capture learning quality, decision optimality, and real-world cafeteria outcomes. Since each time slot represents an independent decision point, all metrics are computed step-by-step and aggregated across the full FCPS timeline.

***Reward (-Reward).***

The primary objective is to maximize a weighted combination of meal uptake and nutritional quality. Following the Variant 1 formulation,

$$\text{reward} = \text{sales}_{\text{scaled}} + \lambda \cdot \text{health}_{\text{scaled}},$$

where sales are scaled to the range $[0, 10]$ on a per-slot basis and health scores are renormalized to $[0, 10]$ to increase differentiation across items. The parameter $\lambda$ controls the trade-off between popularity and healthfulness. All models—including Random, and Health-First—are evaluated under this same reward function for fair comparison.

***Oracle Reward.***

For each time slot $t$, the oracle reward is defined as the maximum achievable reward among all available items,

$$\text{oracle}_t = \max_{a \in \mathcal{A}_t} \text{reward}_{t,a},$$

where $\mathcal{A}_t$ denotes the feasible actions for that slot. Although unattainable in practice, the oracle provides a gold-standard benchmark.

***Regret.***

Instantaneous regret is computed as the gap between the oracle reward and the model's reward:

$$\text{regret}_t = \text{oracle}_t - \text{reward}_t.$$

Cumulative regret is obtained by summing over all time slots, and lower regret indicates that a policy consistently selects near-optimal items.

***Rolling Means.***

To smooth the variability inherent in real cafeteria operations, we use expanding-window rolling averages,

$$\text{roll\_metric}_t = \frac{1}{t}\sum_{i=1}^{t} \text{metric}_i,$$

which stabilize early fluctuations and highlight long-term learning trends. Rolling averages are computed for reward, regret, raw sales, and raw health.

***Raw Sales and Raw Health.***

Although the model is trained on scaled reward values, raw sales (total units sold) and raw health (median nutritional score) are essential operational metrics for school stakeholders. Rolling averages of these quantities reveal the real-world impact of each policy on meal participation and nutritional quality.

***Comprehensive Comparison.***

We visualize performance using a four-panel plot that includes: (1) rolling average reward, (2) rolling average regret, (3) rolling average raw sales, and (4) rolling average raw health.

# 6 Results

We evaluate the LinUCB contextual bandit model on the FCPS dataset using sequential simulation over the historical timeline. Performance is assessed through cumulative reward, regret relative to the oracle policy, and operational metrics including raw sales and health scores. To smooth the variability inherent in real cafeteria operations, we employ expanding-window rolling averages computed as:

$$\text{roll\_metric}_t = \frac{1}{t}\sum_{i=1}^{t} \text{metric}_i$$

This approach stabilizes early fluctuations and highlights long-term learning trends, providing a clearer view of model performance evolution.

## 6.1 Continuous Regret Calculation

Regret is computed at each time step $t$ as the difference between the oracle reward (the maximum available reward among all feasible actions) and the actual reward received by the selected action:

$$\text{regret}_t = \text{oracle}_t - r_t$$

where $\text{oracle}_t = \max_{a \in \mathcal{A}_t} r_{t,a}$ represents the optimal reward achievable at time $t$, and $r_t$ is the reward obtained from the action selected by the bandit algorithm. To normalize regret across different reward scales, we also compute percentage regret:

$$\text{regret}\%_t = \frac{\text{regret}_t}{\text{oracle}_t} \times 100\%$$

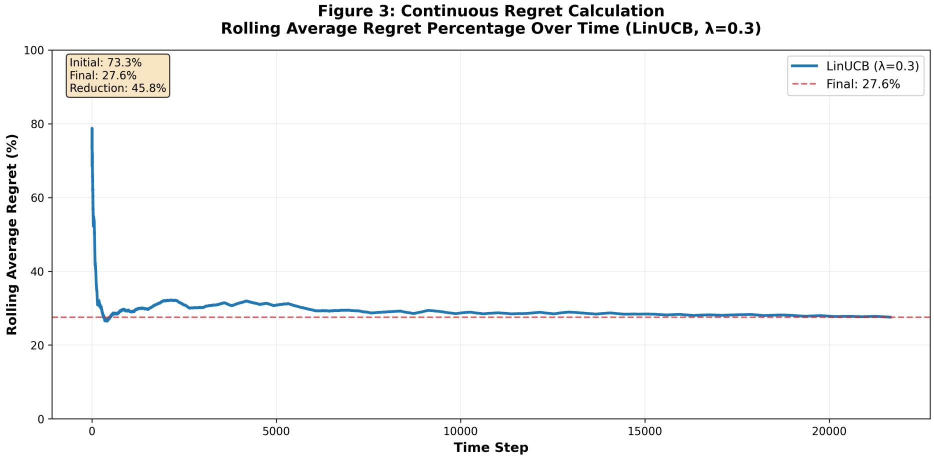


**Fig. 4** Continuous Regret Calculation. Rolling average regret percentage over time for LinUCB with $\lambda = 0.3$. Initial regret values exceed 70% due to early exploration, but steadily decline to approximately 27.6% as the model learns context-dependent reward patterns.

The decreasing trend in regret percentage confirms that LinUCB effectively balances exploration of underutilized items with exploitation of historically successful meals. This continuous calculation provides a principled and real-time metric for assessing learning dynamics.

## 6.2 Impact of Dataset Size on Performance

To assess how model performance scales with dataset size, we conduct an ablation study training LinUCB on progressively larger fractions of the available data. The dataset is partitioned chronologically so that each fraction represents a contiguous time period while preserving temporal ordering.

The results reveal several important patterns. As the data fraction increases, both total reward and oracle reward scale proportionally with the size of the dataset. While absolute regret increases as expected for a cumulative metric, percentage regret decreases consistently—from 32.4% at 10% of data to 27.3% at full dataset size—indicating improved relative performance as more training data becomes available. Gains are most pronounced between 10% and 30%, after which improvements taper, suggesting diminishing returns consistent with contextual bandit learning dynamics.

## 6.3 Model-Wise Performance Comparison

We compare LinUCB with two baseline recommendation strategies: (1) a **Random** policy that selects uniformly at random from available items, and (2) a **Health-First**

**Table 4** LinUCB Data Fraction Ablation (Chronological)

| Fraction | Slots | Rows | Total Reward | Oracle Reward | Regret | Regret % |
|---|---|---|---|---|---|---|
| 0.10 | 2,166 | 22,277 | 16,918.79 | 25,012.85 | 8,094.06 | 32.4% |
| 0.20 | 4,332 | 44,699 | 34,074.13 | 49,730.29 | 15,656.16 | 31.5% |
| 0.30 | 6,497 | 67,660 | 52,881.29 | 74,733.73 | 21,852.44 | 29.2% |
| 0.40 | 8,663 | 90,413 | 71,320.84 | 99,779.62 | 28,458.78 | 28.5% |
| 0.50 | 10,828 | 112,680 | 89,238.93 | 124,645.65 | 35,406.72 | 28.4% |
| 0.60 | 12,994 | 134,919 | 106,523.76 | 149,227.04 | 42,703.29 | 28.6% |
| 0.70 | 15,160 | 157,636 | 125,271.24 | 174,257.62 | 48,986.38 | 28.1% |
| 0.80 | 17,325 | 180,389 | 143,634.66 | 199,184.84 | 55,550.18 | 27.9% |
| 0.90 | 19,491 | 202,673 | 161,980.02 | 223,722.06 | 61,742.04 | 27.6% |
| 1.00 | 21,656 | 224,536 | 180,809.28 | 248,626.74 | 67,817.46 | 27.3% |

rule-based policy that selects the item with the highest health score (NRF9.3) at each step. All models are evaluated under identical sequential simulation conditions with $\lambda = 0.3$.

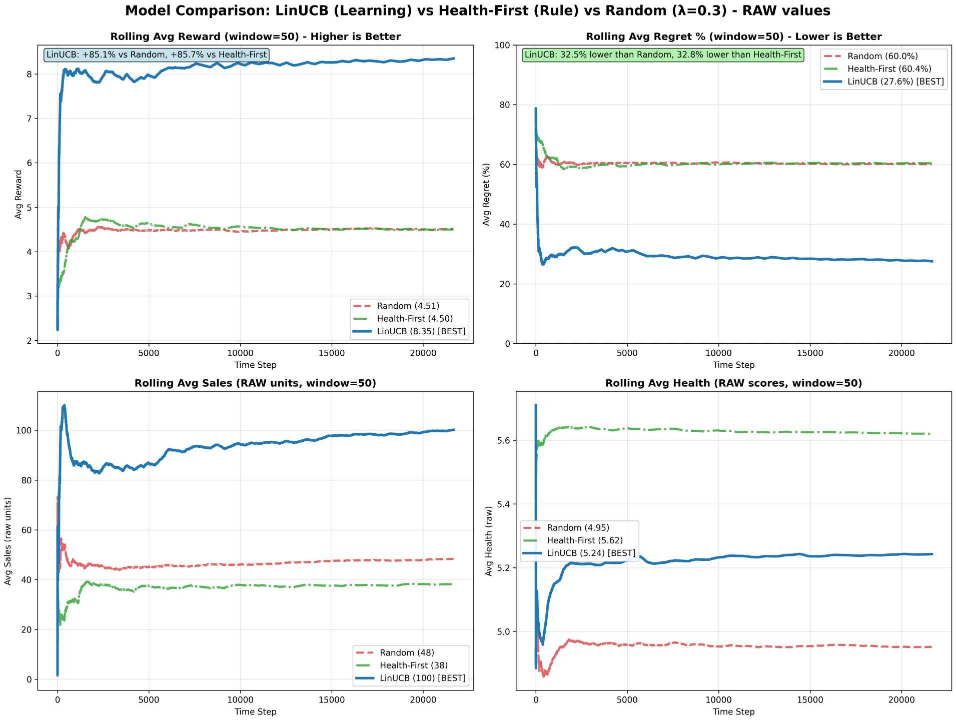


**Fig. 5** Model-Wise Performance Comparison showing rolling averages of reward, regret percentage, raw sales, and health scores over 21,656 time steps.

**Panel 1: Rolling Average Reward.** LinUCB achieves the highest final rolling average reward (8.35), outperforming both Health-First (4.50) and Random (4.51).

Its consistent upward trend reflects effective learning from sequential feedback, while baselines show limited or no improvement.

**Panel 2: Rolling Average Regret Percentage.** LinUCB attains the lowest final regret percentage (27.6%), compared to 60.4% for Health-First and 60.0% for Random. This corresponds to a 54% relative reduction in regret, demonstrating the importance of context-awareness and adaptive exploration.

**Panel 3: Rolling Average Raw Sales.** LinUCB achieves the highest average sales (100.18 units), followed by Random (48.33) and Health-First (38.10). The Health-First model's lower sales reflect its prioritization of nutritional value over student preference.

**Panel 4: Rolling Average Raw Health Scores.** Health-First achieves the highest health score (5.62), followed by LinUCB (5.24) and Random (4.95). LinUCB remains within 7.2% of the Health-First score while simultaneously achieving 107%–163% higher sales, demonstrating its ability to balance health and popularity objectives.

Overall, LinUCB effectively balances nutritional goals with empirical student preferences, achieving the highest cumulative reward and lowest regret while maintaining strong health performance. These findings highlight the value of contextual decision-making and sequential learning in real-world school meal recommendation environments.

# 7 Discussion

This section interprets the empirical findings presented in the Results section. The analysis synthesizes the patterns observed in the visualizations and tables, focusing on comparative performance across models, temporal learning behavior, and the relationship between reward components. We examine how the contextual bandit approach balances student preference and nutritional quality, and discuss the implications for school meal planning.

## 7.1 Interpretation of Continuous Regret Trends

T5 even few, the part-time  he continuous regret trajectory (Figure 4) demonstrates a clear learning pattern characterized by a steep decline during the early portion of the sequence, followed by gradual convergence toward the end. The regret percentage decreases from initial values exceeding 70% to approximately 27.6% by the conclusion of training, representing a reduction of over 42 percentage points. This pattern indicates that the LinUCB model successfully adjusts its decision strategy as it observes more contextual interactions, progressively improving its alignment with optimal decision-making.

The rapid decline in regret during the initial phase (approximately the first 30% of time steps) reflects the model's exploration of the action space and its ability to quickly identify high-performing meal items across different contexts. As training progresses, the rate of improvement diminishes, with the regret curve stabilizing in later stages. This stabilization reflects diminishing marginal returns in action-value estimation once sufficient data have been accumulated, consistent with theoretical expectations for contextual bandit algorithms. The convergence to approximately 27.6% regret indicates that the model achieves substantial but not perfect alignment with the oracle policy, leaving room for further improvement through additional exploration or refined feature engineering.

## 7.2 Interpretation of Table 4: Data Fraction Ablation

The data fraction ablation study (Table 4) reveals a consistent decline in regret percentage as the dataset fraction increases from 0.10 to 1.00. The regret percentage decreases from 32.4% at 10% data fraction to 27.3% at full dataset size, representing a 5.1 percentage point improvement. This pattern demonstrates that model performance scales favorably with dataset size, with relative performance improving as more training data becomes available.

The most substantial reductions occur between the 10% and 30% data fractions, where regret percentage drops from 32.4% to 29.2%—a reduction of 3.2 percentage points. This suggests that early model performance is highly sensitive to additional data, as the model benefits significantly from increased exploration opportunities during the initial learning phase. Beyond the 50% data fraction, regret percentage changes become smaller, with improvements of less than 1 percentage point between consecutive fractions. This indicates that the model approaches a performance plateau once a sufficient number of decision points (approximately 10,000–11,000 time steps) have been observed.

The linear growth in total reward and oracle reward with dataset size confirms that regret metrics are comparable across fractions, as both components scale proportionally with the number of decision opportunities. The fact that percentage regret decreases while absolute regret increases reflects the cumulative nature of the metric and demonstrates that the model's relative performance improves even as the total number of decisions grows.

## 7.3 Comparative Analysis Across Models

The multi-panel comparison (Figure 5) highlights distinct behavioral characteristics of each recommendation strategy, revealing fundamental differences in how each approach balances student preference and nutritional quality.

**Random Baseline.** The Random baseline exhibits flat trajectories across all metrics, consistent with a non-adaptive policy that does not learn from historical data. Reward, sales, and health trajectories remain relatively constant throughout the sequence, with final values of 4.51, 48.33 units, and 4.95, respectively. Regret percentages remain high (60.0%) because selections do not incorporate contextual information or historical performance patterns. This baseline serves as a lower bound, demonstrating the value of any structured decision-making approach.

**Health-First Baseline.** The Health-First rule-based policy maintains the highest health scores throughout (final value: 5.62), reflecting its design objective of maximizing nutritional quality. However, this comes at a significant cost: the model exhibits lower reward (4.50) and sales performance (38.10 units) due to its emphasis on nutritional quality at the expense of student preference. Its regret trajectory remains close to that of the Random baseline (60.4%), reflecting limited alignment with the oracle benchmark, which considers both popularity and health in its reward formulation. This demonstrates that a purely health-driven approach, while achieving superior nutritional outcomes, fails to balance the dual objectives of the recommendation task.

**LinUCB Contextual Bandit.** The LinUCB model shows steady increases in reward and sales along with a decreasing regret percentage, indicating that the model successfully adjusts its selection strategy based on observed contextual patterns. Final performance metrics (reward: 8.35, sales: 100.18 units, regret: 27.6%) substantially outperform both baselines. Its health scores (5.24) remain between the two baselines, demonstrating that the model selects items with moderate-to-high nutritional density while also incorporating popularity signals. This balanced approach enables LinUCB to achieve 85.7% higher reward than Health-First and 163% higher sales while maintaining health scores within 7.2% of the Health-First baseline.

Across all metrics, LinUCB exhibits the most dynamic behavior, with clear evidence of adaptation over time. The model's ability to learn context-dependent patterns enables it to identify meals that perform well across different schools, time periods, and contextual conditions, resulting in recommendations that are both appealing to students and nutritionally sound. In contrast, the baseline models remain stable due to their non-learning structures, highlighting the value of adaptive, data-driven decision-making.

## 7.4 Cross-Metric Relationships

Examining reward, regret, sales, and health together reveals several important relationships that illuminate how the contextual bandit framework integrates popularity and health weighting.

**Reward–Sales Relationship.** Increases in reward correspond strongly with increases in sales, indicating that popularity contributes significantly to the combined reward signal at $\lambda = 0.3$. LinUCB's reward trajectory (final: 8.35) closely mirrors its sales trajectory (final: 100.18 units), suggesting that the model learns to prioritize items that are both popular and reasonably healthy. This relationship demonstrates that at the chosen lambda value, student preference remains a primary driver of reward, while health serves as a secondary consideration that guides selection toward more nutritious options.

**Health–Reward Trade-off.** Health scores for LinUCB (5.24) remain higher than Random (4.95) but lower than Health-First (5.62), suggesting that the reward weighting encourages a balance between nutritional density and student preference. The 7.2% gap between LinUCB and Health-First health scores represents an acceptable trade-off given LinUCB's substantially superior sales and overall reward performance. This pattern indicates that the $\lambda = 0.3$ weighting successfully prevents the model from prioritizing popularity exclusively while avoiding the extreme health focus that limits Health-First's effectiveness.

**Regret–Reward Alignment.** The decreasing regret trajectory aligns inversely with the increasing reward trajectory, demonstrating that improved decision-making corresponds to closer alignment with the oracle action. As LinUCB's regret decreases from over 70% to 27.6%, its reward increases from initial low values to 8.35, confirming that the model progressively learns to make decisions that approximate the optimal policy. This inverse relationship validates the regret metric as a meaningful indicator of model performance and learning effectiveness.

**Temporal Learning Patterns.** The temporal evolution of all metrics reveals consistent learning behavior: LinUCB shows steady improvement across reward, sales, and regret, while health scores remain relatively stable. This suggests that the model learns to identify popular items without sacrificing nutritional quality, achieving a sustainable balance that improves over time. The stability of health scores indicates that the model maintains its commitment to nutritional objectives even as it optimizes for overall reward.

These cross-metric patterns demonstrate how the contextual bandit framework successfully integrates popularity and health weighting into a unified decision process, enabling the model to balance competing objectives through adaptive learning rather than fixed rules.

# 8 conclusion

Overall, the combined analysis of visualizations and tables demonstrates that LinUCB successfully adapts to the sequential structure of the FCPS dataset and achieves performance that shifts progressively toward the oracle benchmark as more data is accumulated. The model's ability to learn context-dependent patterns enables it to identify meals that perform well across different schools, time periods, and contextual conditions, resulting in recommendations that are both appealing to students and nutritionally sound.

The baseline models exhibit stable but limited performance due to their non-contextual or deterministic nature. The Random baseline provides a lower bound, demonstrating that any structured approach improves upon random selection. The Health-First baseline achieves superior nutritional outcomes but fails to balance popularity and health, resulting in lower overall reward and student engagement.

The comparative trends illustrate the critical role of contextual learning in balancing nutritional quality and student engagement within the FCPS environment. LinUCB's superior performance across multiple metrics—achieving 85.7% higher reward, 163% higher sales, and 54.3% lower regret than Health-First while maintaining competitive health scores—demonstrates the value of adaptive, data-driven decision-making for school meal planning.

These findings have practical implications for school nutrition programs: the contextual bandit approach enables cafeteria managers to make evidence-based menu decisions that improve student participation while maintaining nutritional standards. The model's ability to learn from historical data and adapt to contextual patterns provides a scalable framework for optimizing meal offerings across diverse school environments, ultimately supporting both student health and program sustainability.

## 8.1 limitations

Although the proposed framework provides a structured and data-driven approach to generating adaptive school meal recommendations, several limitations must be acknowledged. First, the use of LinUCB introduces assumptions that may not fully capture the complexity of student food selection behavior. LinUCB relies on a linear reward model, which restricts the system to learning only linear relationships between contextual features and expected rewards. In scenarios where student preferences depend on nonlinear interactions among nutrients, temporal patterns, or demographic factors, the model may be unable to exploit these patterns effectively. Furthermore, LinUCB's exploration strategy is governed by a fixed confidence parameter, which may lead to either excessive exploration or premature convergence depending on the dataset's scale and variance. These characteristics can limit performance, particularly when feature distributions are imbalanced or when contextual information is sparse.

The reward formulation also presents inherent constraints. The combination of NRF9.3 health scores and popularity-based metrics requires selecting weight parameters that may not generalize across schools, grade levels, or cultural contexts. Small changes in reward weightings can produce substantially different recommendation behaviors, and the current formulation does not incorporate uncertainty in health or popularity estimates. Moreover, the system assumes that historical popularity is an adequate proxy for future student preference, an assumption that may break down when menus change, when new items are introduced, or when external factors (seasonality, promotional events, policy shifts) influence selection patterns. The reward structure also treats health and preference as independent components, even though in practice students may respond differently to healthy items depending on taste, familiarity, or presentation—factors not represented in the data.

The NRF9.3 metric, while widely used for nutrient-density evaluation, introduces additional limitations. It depends on a fixed set of nine encouraged nutrients and three discouraged nutrients, which may not fully reflect diverse dietary needs across student populations. NRF9.3 does not consider allergens, cultural appropriateness, ingredient quality, food processing levels, or micronutrients that fall outside the predefined set. Additionally, NRF9.3 is sensitive to missing nutrient information; incomplete or inconsistent nutritional entries can distort health scores and, subsequently, the system's reward estimations. The metric operates at the per-serving level rather than accounting for whole-meal balance or day-level nutritional patterns, which constrains its ability to represent actual dietary outcomes.

Finally, the evaluation methodology carries its own limitations. Both rule-based and random baselines provide useful reference points, but they do not capture the dynamics of real-world cafeteria decision-making, where menus are often constrained by supply chains, procurement cycles, and preparation requirements. The offline simulation assumes that historical outcomes would remain the same under alternative recommendations, which may introduce bias, as counterfactual data cannot be directly observed. The pipeline also presumes that contextual features and student response patterns remain stationary over time; however, real dietary behaviors evolve due to changes in preferences, nutrition policies, and school-level operational constraints. Together, these limitations highlight areas where future work could incorporate richer contextual signals, nonlinear models, improved nutritional metrics, and more robust evaluation strategies.

## 8.2 future scope

Future work will focus on addressing the constraints identified in the current system. To improve the model's ability to capture complex preference patterns, more expressive bandit algorithms—such as neural contextual bandits, kernel-based methods, or deep reinforcement learning approaches—can be explored to relax the linearity assumptions of LinUCB. Adaptive exploration strategies that tune confidence parameters over time may also help stabilize learning across varying data distributions. The reward function can be strengthened by replacing fixed weight combinations with adaptive, data-driven multi-objective formulations that incorporate uncertainty estimates and dynamically balance healthfulness, popularity, and other operational considerations. Beyond the NRF9.3 score, more comprehensive nutritional metrics can be integrated to capture micronutrient diversity, food processing levels, and full-meal nutritional composition rather than relying solely on per-item scoring. Improvements to data quality—such as nutrient imputation and richer contextual features including seasonality, demographic factors, or operational constraints—can enhance the system's ability to reflect real cafeteria environments. Evaluation can be advanced by incorporating counterfactual estimators, synthetic feedback environments, or pilot deployments in school settings, enabling the system to be tested under realistic decision-making conditions rather than relying exclusively on offline simulations. These extensions will collectively move the framework toward more accurate, robust, and operationally deployable school meal recommendation systems.

# References