



The George Washington University

Data Science Program

**Forecasting School Meal Production Costs:  
A Comparative Study of Machine Learning and Deep  
Learning  
Time-Series Models**

Varshith Reddy

Areena

Chaya

Capstone Project Report

Fall 2025

## **Abstract**

Meal production planning in Fairfax County Public Schools (FCPS) is difficult due to irregular student participation and fluctuating demands for meals, resulting in a high percentage of food waste. These patterns are highly nonlinear and difficult to model with conventional forecasting approaches, making it a significant challenge for school administrators. In this project we designed, implemented, and compared several machine learning and deep learning models, including Linear Regression, XGBoost, Feed-Forward Neural Networks (FNN), Gated Recurrent Units (GRU), and Long Short-Term Memory (LSTM) networks using real FCPS multivariate meal-production data. Among all models, the univariate LSTM achieved the best performance, reducing the mean forecasting error from about 197 to 137 (30.6% improvement). These results show that deep learning methods can substantially improve production cost forecasting, support more accurate budgeting, reduce food waste, and enable better data-driven decision making in school meal programs.

# Contents

<b>1</b>	<b>Introduction</b>	<b>4</b>
<b>2</b>	<b>Problem Statement</b>	<b>5</b>
<b>3</b>	<b>Related Work</b>	<b>5</b>
3.1	Food Waste in School Cafeterias . . . . .	5
3.2	Forecasting Models and Deep Learning . . . . .	5
<b>4</b>	<b>Solution and Methodology</b>	<b>6</b>
4.1	Data Acquisition and Parsing Pipeline . . . . .	6
4.2	Exploratory Data Analysis (EDA) . . . . .	8
<b>5</b>	<b>Results and Discussion</b>	<b>14</b>
5.1	Experimentation Protocol . . . . .	14
5.2	Data Tables: . . . . .	15
5.3	Graphs . . . . .	15
5.3.1	Multivariate Model Plots : . . . . .	16
5.3.2	Uniivariate Model Plots : . . . . .	19
5.3.3	Forecasting Example . . . . .	21
<b>6</b>	<b>Discussion</b>	<b>22</b>
<b>7</b>	<b>Conclusion</b>	<b>23</b>

## List of Figures

1	Correlation heatmap showing pairwise Pearson correlations among numerical features.	9
2	Singular value magnitudes from the SVD of the standardized feature matrix. . . . .	10
3	Time series and forecasting pipeline. . . . .	12
4	Multivariate GRU Predicted vs Actual . . . . .	16
5	Multivariate LSTM Predicted vs Actual . . . . .	16
6	Multivariate FNN Predicted vs Actual . . . . .	17
7	Multivariate Linear Regression Predicted vs Actual . . . . .	17
8	Multivariate XGBoost: Predicted vs Actual . . . . .	18
9	Univariate GRU Predicted vs Actual . . . . .	19
10	Univariate LSTM Predicted vs Actual . . . . .	19
11	Univariate FNN Predicted vs Actual . . . . .	20
12	Univariate XGBoost Predicted vs Actual . . . . .	20
13	Univariate Linear Regression Predicted vs Actual . . . . .	21
14	Univariate LSTM Train/Test/Forecast Example . . . . .	21

## List of Tables

1	Data dictionary of input features. . . . .	7
2	Columns Used for Forecasting Models . . . . .	8
3	Multivariate Model Performance . . . . .	15
4	Univariate Model Performance . . . . .	15

# 1 Introduction

Forecasting daily production costs in school meal programs is one of the most enduring challenges faced by district administrators, nutrition service directors, and operational planners. Even with the enhancement of digital record-keeping and standardized menus in schools, there are major fluctuations in areas such as student participation, ingredient prices, and food waste. These fluctuations complicate budgeting, disrupt daily meal planning, and increase operational inefficiencies. Traditional approaches to forecasting—which typically include heuristic rules or a linear extrapolation—cannot capture nonlinear, irregular patterns, and time dependencies present in real production-cost data, at least as influenced by school-specific behaviors, holidays, and supply-chain variability.

Recent developments in machine learning and deep learning allow models to learn complex temporal dependencies directly from their historical data. In this work, we develop an integrated framework in forecasting short-term production costs by using a dataset created through scraping and merging over 100 HTML production reports of various FCPS schools. This resulted in challenges such as inconsistent currency formatting, irregular date structures, missing entries, and extreme outliers; these are dealt with through a comprehensive preprocessing pipeline that involves currency coercion, outlier removal, automatic date parsing, temporal aggregation, and sliding-window sequence generation.

With the cleaned dataset, we assess the performance of various models, including Linear Regression, XGBoost, Feed-Forward Neural Networks, Gated Recurrent Units, and Long Short-Term Memory networks for both univariate and multivariate settings. All models are trained on a 70-30 train-test split and tested on the test set using standard error-based metrics: MSE, RMSE, MAE, and  $R^2$ . Several diagnostic plots, such as predicted-vs-actual graphs and multi-day forecasts, have also been created to assess each model’s performance in terms of capturing the underlying temporal structure and variability in the costs.

The aim of this work is to provide a broad comparison of forecasting techniques for school meal production-cost forecasting and to offer a practical framework that districts can use to improve budget planning, reduce food waste, and enhance overall operational efficiency. The sections that follow outline the dataset and preprocessing (Section 2), exploratory analysis (Section 3), methodology and models (Section 4), results (Section 5), practical implications (Section 6), and conclusions with future directions (Section 7).

## 2 Problem Statement

Daily production cost forecasting in FCPS school meal programs is a challenging task as real cost values change unpredictably due to menu differences, student participation, leftover amounts, discarded food, and ingredient price fluctuations. Collected raw data from FCPS, which consists of over 100 HTML production reports, includes inconsistent currency format representations, missing and irregular entries, school-specific variations, and extreme outliers, further complicating the forecasting problem. These nonlinear, time-varying patterns cannot be reflected by manual estimates or historical averages on which FCPS relies for planning.

This project’s challenge is to transform messy, unstructured FCPS production reports into a clean and reliable time-series dataset and determine which forecasting approach predicts next-day production costs most accurately. After creating `meals_combined.csv` using HTML scraping, preprocessing, outlier removal, and sliding-window generation, we implemented the following machine learning and deep learning models in univariate and multivariate settings: Linear Regression, Feed-Forward Neural Networks, XGBoost, GRU, and LSTM. The goal is to determine what type of model learns temporal cost patterns most effectively so FCPS can reduce overproduction, lower food waste, and better plan budgets.

## 3 Related Work

Research into the management of school meals falls within two broad categories: studies on food waste in cafeterias and studies on methods of forecasting. Together, these give a background on what exactly this project is trying to address and the limitations of existing approaches.

### 3.1 Food Waste in School Cafeterias

Food waste in school cafeterias has been documented as a financial and environmental issue. Cohen et al. (2013) found that students wasted nearly 20

In fact, after the introduction of healthier meal options, such as fruits and vegetables, Byker et al. (2014) found that food waste increased by 56

Smith and Cunningham-Sabo (2014) showed that waste rates vary greatly across menu items; some items were wasted at a rate over 35

Combined, these studies underscore the fact that poor predictions of meal demand lead to food and financial waste; thus, motivating the need for better forecasting methods to anticipate production needs before meal preparation.

### 3.2 Forecasting Models and Deep Learning

Complementary to food-waste research, improvements in time-series forecasting-most recently with deep learning-offer powerful instruments to model complex patterns. Hochreiter and Schmidhuber 1997 proposed Long Short-Term Memory (LSTM) networks, which are designed to handle the

disadvantage of previous RNNs by keeping long-term information using their gating mechanism. Later, Cho et al. 2014 developed the Gated Recurrent Unit (GRU), a faster alternative to LSTMs. Chung et al. 2014 showed that LSTMs generally provide better performance than GRUs on sequences of longer length, but GRUs performed roughly as well on shorter forecasting horizons, as seen in the 7-day predictions in this study.

Deep learning has returned promising results in many sectors. Kong et al. (2019) used LSTMs to forecast residential electricity consumption and achieved a low error rate of only 2.28

Despite these developments, no previous studies have used deep learning or state-of-the-art forecasting approaches to forecast school meal production costs. The few school-related papers only measure post-meal wastes, while the forecasting papers target industries like energy, retail, and healthcare rather than school operations.

Taken together, food-waste studies reveal the consequence of poor meal planning, while forecasting research demonstrates that state-of-the-art sequence-learning models can effectively learn nonlinear and time-varying patterns. However, these research streams are separated. Food-waste studies provide no predictive solution, whereas forecasting studies fail to consider the operational challenges faced by school meal production.

This project bridges that gap by applying machine learning and deep learning models, including Linear Regression, XGBoost, Feed-Forward Neural Networks, GRU, and LSTM, to real production-cost data scraped from over 100 HTML reports. Unlike any other studies before them, our study focuses explicitly on forecasting production costs prior to preparation, providing a data-driven framework through which school districts can seek to minimize waste, enhance the accuracy of budgeting, and inform more efficient meal-planning decisions.

## 4 Solution and Methodology

This section details the overall designed forecasting solution for this project, including the data acquisition pipeline, preprocessing procedures, exploratory analysis, time-series construction, model architectures, evaluation metrics, and backtesting strategy.

### 4.1 Data Acquisition and Parsing Pipeline

This study was based on de-identified, publicly available daily production reports published from FCPS. The raw data did not come in a tabular form; rather, it was available in over 100 HTML files, each containing item-level tables at breakfast and lunch across more than 100 schools. These tables included quantities planned, served, discarded, leftover, and the total production cost.

To transform these unstructured reports into an analysis-ready dataset, a custom HTML parsing pipeline was implemented. This pipeline automatically:

- detected school headers and meal sections
- Extracted item-level tables
- removed subtotal rows

standardized column names

cleaned numeric fields, including currency, integers, percentages

Corrected missing entries by using forward/backward fill

Removed invalid or empty rows.

Currency strings like “\$13.32” were converted to numeric values, symbols and commas were removed; and percentage fields were converted into floating-point values. After processing all HTML files, breakfast and lunch datasets were combined into a single structured dataset called meals\_combined.csv, with approximately 177,492 records and 20 columns.

Table 1: Data dictionary of input features.

Name of Feature	Description of Feature
school_name	Name of the FCPS school associated with the production report.
date	Calendar date of the meal production record (YYYY-MM-DD).
identifier	Unique internal identifier for each item entry.
name	Name of the meal item (e.g., entrée, milk, side dish).
planned_reimbursable	Number of planned reimbursable meals.
planned_non_reimbursable	Number of planned non-reimbursable meals.
planned_total	Total number of meals planned.
offered_total	Total quantity of items offered.
served_reimbursable	Count of reimbursable meals served.
served_non_reimbursable	Count of non-reimbursable meals served.
served_total	Total meals served.
served_cost	Cost associated with served items.
discarded_total	Number of meals discarded.
discarded_percent_of_offered	Percentage of offered items discarded.
discarded_cost	Cost of discarded meals.
subtotal_cost	Intermediate cost measure.
left_over_total	Number of leftover meals.
left_over_percent_of_offered	Percentage of offered items left over.
left_over_cost	Cost associated with leftover items.
production_cost_total	Total production cost (target variable).
meal_type	Meal category (Breakfast or Lunch).



Table 2: Columns Used for Forecasting Models

Column	Description
Date	Time index for forecasting sequences.
School Name	Grouping variable maintaining school-specific continuity.
Meal Type	Breakfast or lunch identifier.
Served Total	Lagged feature capturing consumption behavior.
Planned Total	Planned demand indicator.
Discarded Total	Waste-related signal contributing to variability.
Left Over Total	Operational efficiency indicator.
Production Cost Total	Primary target variable for forecasting tasks.

## 4.2 Exploratory Data Analysis (EDA)

Extensive EDA was performed to understand the statistical properties of the dataset, assess data quality, verify distributional behavior of key variables, and check for potential multicollinearity among predictors. The main focuses of interest were the *production\_cost\_total* variable, operational meal counts, and structural relationships among these features. All analyses were conducted after the dataset was fully cleaned and standardized.

### Outlier Detection and Preprocessing

During the preliminary analysis, the *production\_cost\_total* variable exhibited a strong right-skew due to rare extreme anomalies. One such anomaly appeared in the raw dataset: the item “*Fat Free Chocolate Milk*” at Hughes Middle School showed a production cost of \$14,010 for a single day. This value, though real in the dataset, is implausible in practice and justified the need for strict outlier removal to stabilize model training.

#### Outlier removal used:

- **IQR-based detection:**

A value  $x$  is considered an outlier if:

$$x < Q_1 - 1.5 \times IQR \quad \text{or} \quad x > Q_3 + 1.5 \times IQR.$$

- **99th-percentile thresholding:**

Used only for *production\_cost\_total* to remove extreme spikes.

- Applied only to the target variable to eliminate rare extreme values.
- Helps stabilize the distribution before training.
- Ensures that unrealistic anomalies do not distort model learning.

Whenever possible, missing values have been filled, and rows that could not be logically reconstructed have been removed. Dates were standardized to ensure temporal continuity across all schools and meal types.

## Correlation Analysis

A Pearson correlation matrix (Fig. 1) revealed strong correlation between *served\_total* and *production\_cost\_total*, confirming cost dependence on volume served.

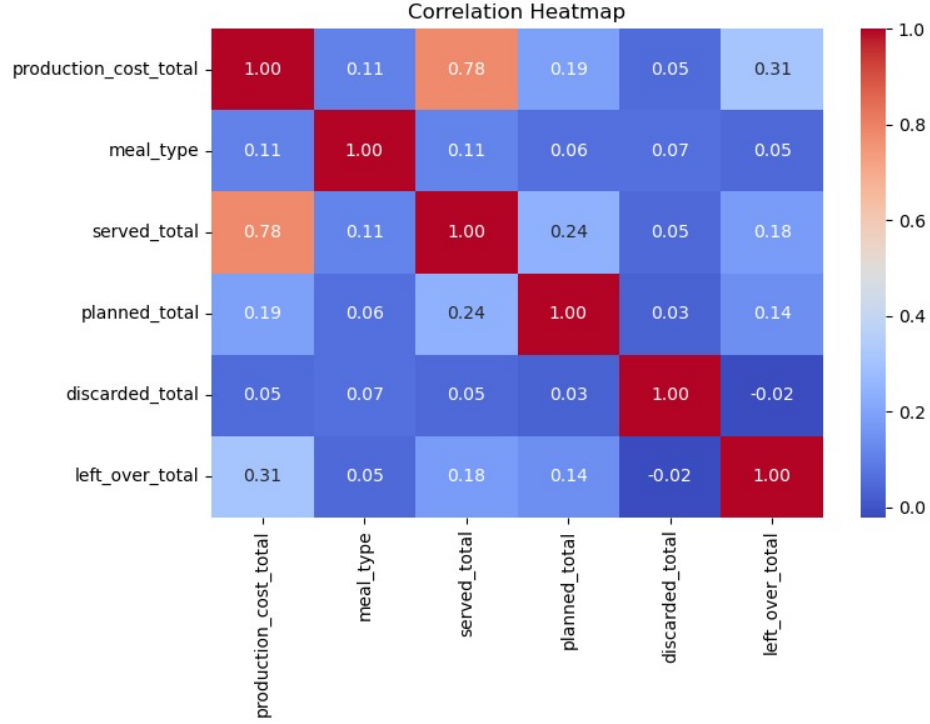


Figure 1: Correlation heatmap showing pairwise Pearson correlations among numerical features.

## Multicollinearity Diagnostics

Two tests were conducted:

- **Variance Inflation Factor (VIF):**

$$\text{VIF}(X_i) = \frac{1}{1 - R_i^2}$$

All VIF values ranged from **1.02–1.28**, indicating negligible multicollinearity.

- **Singular Value Decomposition (SVD):**

$$X = U\Sigma V^T$$

All singular values were considerably above zero, which means the dataset is well-conditioned.

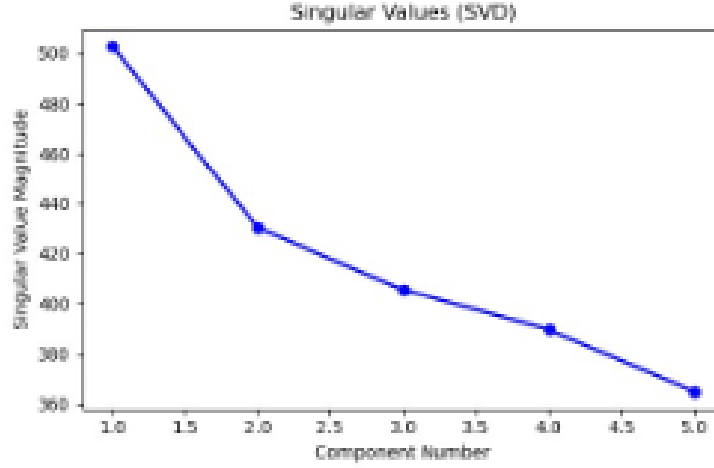


Figure 2: Singular value magnitudes from the SVD of the standardized feature matrix.

Overall, EDA validated that the cleaned dataset was amenable for downstream time-series modeling.

### Time-Series Construction

For each school  $s$  and meal type  $m$ , an individual daily time series  $\{y_t^{(s,m)}\}$  was created by aggregating all item-level *production\_cost\_total* values for that day.

A sliding-window approach was used to create supervised learning samples. For a window size  $W = 3$ :

- **Input:** last 3 days of costs
- **Output:** next-day cost

The forecasting target is expressed as:

$$\hat{y}_{t+1} = f(y_t, y_{t-1}, y_{t-2})$$

Each school-meal pair was modeled independently to preserve school-specific behavior.

### Data Scaling and Splitting

Each time series was normalized using Min-Max scaling:

$$\tilde{y}_t = \frac{y_t - \min(y_{\text{train}})}{\max(y_{\text{train}}) - \min(y_{\text{train}})}$$

Chronological splits were used:

- 60% training
- 20% validation
- 20% testing

This prevented information leakage and preserved temporal ordering.

## Forecasting Pipeline

The full forecasting pipeline Fig. 2 included:

- Load cleaned dataset
- Split into train-validation-test sets
- Scale training series
- Apply sliding-window transformation
- Optional oversampling for deep models
- Train models with early stopping
- Generate predictions on the test set
- Evaluate performance using statistical metrics

It was this modular structure that ensured reproducibility and model comparability.

## Forecasting Models

The following models were evaluated in both multivariate and univariate settings:

- **Linear Regression**
- **XGBoost Regressor**
- **Feedforward Neural Network (FNN)**
- **Gated Recurrent Unit (GRU)**
- **Long Short-Term Memory Network (LSTM)**

**Univariate vs. Multivariate Inputs.** Univariate models used only historical *production\_cost\_total* sequences. Multivariate models used five operational features: *served\_total*, *planned\_total*, *discarded\_total*, *leftover\_total*, and *production\_cost\_total*.

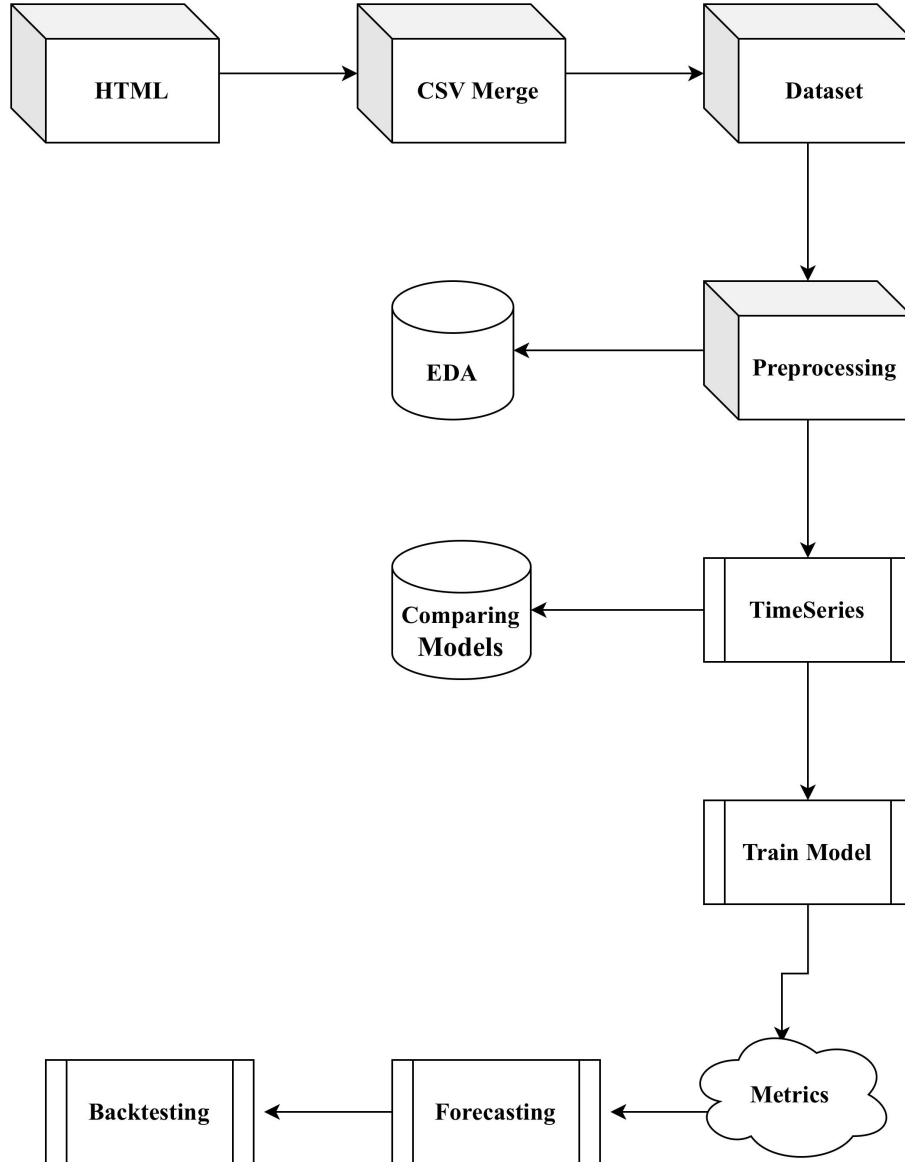


Figure 3: Time series and forecasting pipeline.

**The univariate LSTM outperformed all other models in terms of accuracy and stability.**

### Multivariate Forecasting

In the multivariate setting, each daily observation is represented by a feature vector:

$$\mathbf{x}_t = \begin{bmatrix} \text{served\_total} \\ \text{planned\_total} \\ \text{discarded\_total} \\ \text{leftover\_total} \\ \text{production\_cost\_total} \end{bmatrix} \in \mathbb{R}^5.$$

Given a window length  $W$ , the forecasting task is:

$$\hat{y}_{t+1} = f(\mathbf{x}_t, \mathbf{x}_{t-1}, \dots, \mathbf{x}_{t-W+1}).$$

Here,  $\hat{y}_{t+1}$  is the predicted next-day production cost. Linear Regression, XGBoost, FNN, GRU, and LSTM models were trained on these multivariate sequences to evaluate whether incorporating additional operational variables (served, planned, discarded, leftover) improves accuracy over using the cost alone.

### Univariate Forecasting

In the univariate setting, the model uses only the historical production cost for each school-meal series. For school  $s$  and meal type  $m$ , we construct a time series  $\{y_t^{(s,m)}\}$ , where  $y_t^{(s,m)}$  is the total production cost on day  $t$ .

With window size  $W$ , the forecasting problem becomes:

$$\hat{y}_{t+1}^{(s,m)} = f\left(y_t^{(s,m)}, y_{t-1}^{(s,m)}, \dots, y_{t-W+1}^{(s,m)}\right).$$

The same set of models (Linear Regression, XGBoost, FNN, GRU, and LSTM) was evaluated in this univariate format. In our experiments, the univariate LSTM consistently achieved the lowest error and highest  $R^2$ , so it was selected as the final model for deployment and backtesting.

### Evaluation Metrics

Forecast accuracy was computed using the following metrics:

**Mean Squared Error (MSE):**

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

**Root Mean Squared Error (RMSE):**

$$\text{RMSE} = \sqrt{\text{MSE}}$$

**Mean Absolute Error (MAE):**

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

**Coefficient of Determination ( $R^2$ ):**

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}$$

## Backtesting Procedure

Forecasts were evaluated using a walk-forward backtesting strategy:

- At each test index  $t$ , the model received the last  $W$  true observations.
- Then it predicted the next-day cost:

$$\hat{y}_{t+1}$$

- Predictions were compared against real values using the defined metrics.

This setup replicates real FCPS deployment, where only past data are available when forecasting the next day.

## 5 Results and Discussion

This section presents the performance evaluation of all the forecasting models tested in this paper, together with the experiments conducted to evaluate the accuracy of the proposed forecasting solution. The results are supported with figures, tables, and narrative explanations that interpret the findings.

### 5.1 Experimentation Protocol

All forecasting experiments were conducted on the cleaned `meals_combined.csv` dataset using the daily production cost time series for each school-meal pair.

The workflow was as follows:

- **Dataset Split:**  
Each time series was divided chronologically into 80% training and 20% testing.
- **Model Inputs:**
  - ✓ **Multivariate setting:** Inputs included `served_total`, `planned_total`, `discarded_total`, `leftover_total`, and `production_cost_total`.
  - ✓ **Univariate setting:** Only past values of `production_cost_total` were used.
- **Models Evaluated:**
  - ✓ Linear Regression (LR)
  - ✓ XGBoost
  - ✓ Feedforward Neural Network (FNN)
  - ✓ Gated Recurrent Unit (GRU)
  - ✓ Long Short-Term Memory (LSTM)

- **Evaluation Metrics:**

The following metrics were used for evaluation:

- ✓ Mean Squared Error (MSE)
- ✓ Root Mean Squared Error (RMSE)
- ✓ Mean Absolute Error (MAE)
- ✓ Coefficient of Determination ( $R^2$ )

This protocol ensures a fair comparison across all models and reflects realistic deployment conditions whereby models must predict future costs using only past values.

## 5.2 Data Tables:

Tables 3 and 4 summarize the quantitative results of all multivariate and univariate models.

Table 3: Multivariate Model Performance

Model	MSE	MAE	$R^2$
Linear Regression	662.3026	<b>13.7807</b>	0.4389
XGBoost	506.5291	11.7514	0.5709
FNN	677.3214	12.3852	0.4262
LSTM	<b>259.8</b>	14.2605	<b>0.81</b>
GRU	213.1059	10.3730	0.7734

Table 4: Univariate Model Performance

Model	MSE	RMSE	$R^2$
Linear Regression	39013.54	197.52	0.75
XGBoost	39595.32	198.99	0.74
FNN	39241.72	198.10	0.75
LSTM	<b>18793.44</b>	<b>137.08</b>	<b>0.86</b>
GRU	20186.30	142.07	0.85

## Interpretation:

The tables show that LSTM consistently produced the lowest error in both evaluations; hence, it is the strongest model for school meal production cost forecasting.

## 5.3 Graphs

This section presents visual validation of model performance.



### 5.3.1 Multivariate Model Plots :

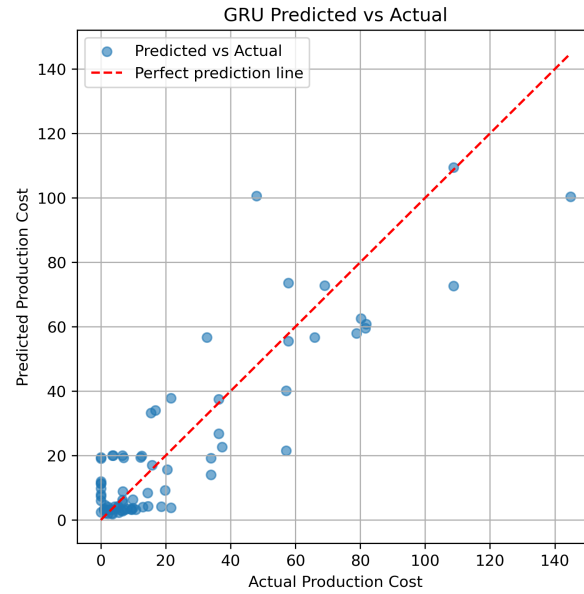


Figure 4: Multivariate GRU Predicted vs Actual

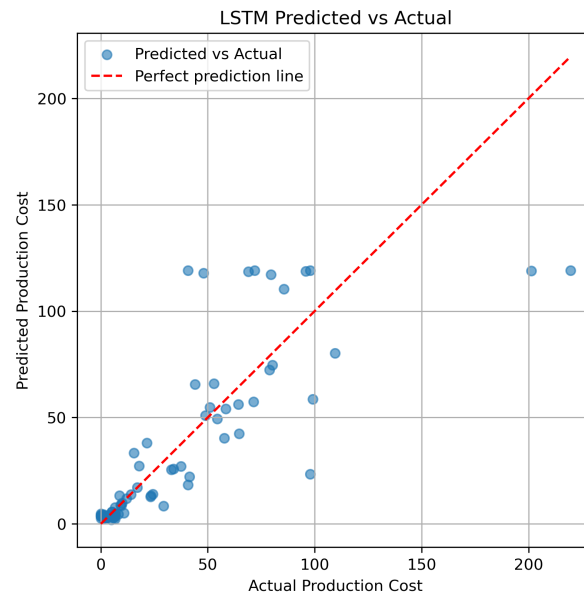


Figure 5: Multivariate LSTM Predicted vs Actual

- The GRU model also performs strongly, with many points near the diagonal. A bit more spread appears for higher values, but overall, GRU makes consistent and reliable predictions using multiple features.
- Most points are close to the diagonal line, showing that the LSTM predicts very accurately

when multiple features are used. It captures the relationship between served meals, planned meals, waste, and cost very well.

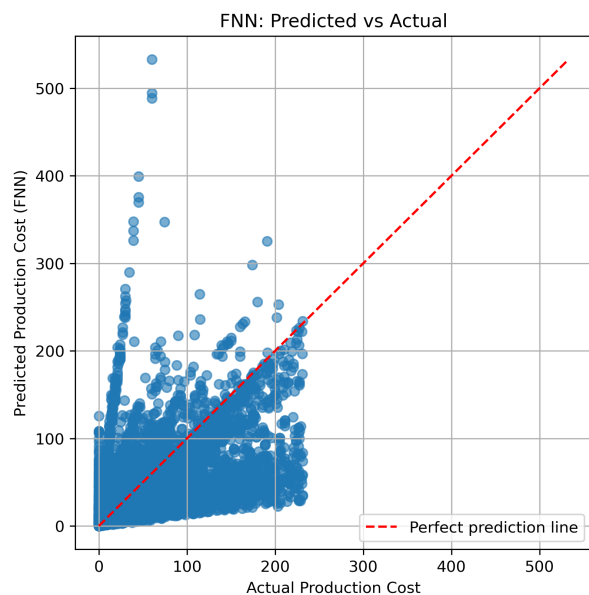


Figure 6: Multivariate FNN Predicted vs Actual

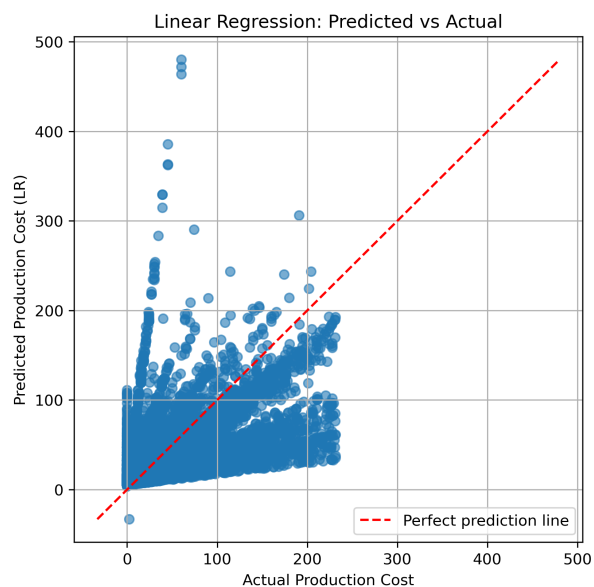


Figure 7: Multivariate Linear Regression Predicted vs Actual

- FNN captures the general direction of the data, but the points are widely spread. This indicates the model struggles with sharp changes in cost and cannot fully use all features to make precise predictions.

- Linear Regression shows a clear trend but has a wide scatter away from the diagonal. This means its predictions are less accurate because a simple linear model cannot fully learn complex patterns in the data.

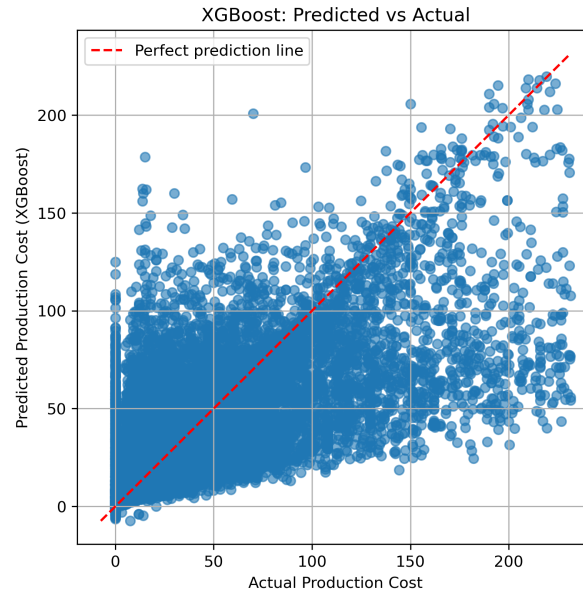


Figure 8: Multivariate XGBoost: Predicted vs Actual

- XGBoost predicts reasonably well, and many points fall near the diagonal line. However, the scatter increases for higher cost values, showing that while it learns non-linear patterns, it sometimes struggles on more extreme days.

### 5.3.2 Uniivariate Model Plots :

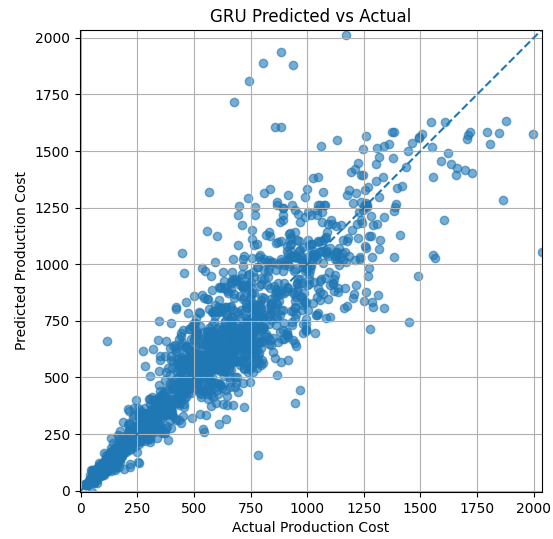


Figure 9: Univariate GRU Predicted vs Actual

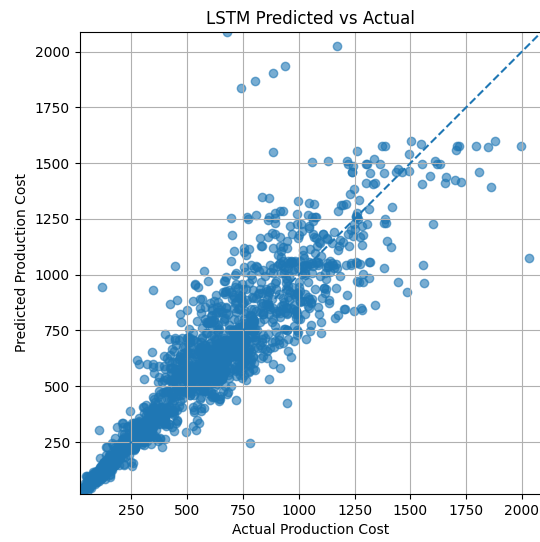


Figure 10: Univariate LSTM Predicted vs Actual

- Most points lie close to the diagonal line, which means LSTM predicts very accurately. It captures both daily changes and longer patterns.
- The GRU points are also near the diagonal, showing good performance. However, the scatter is slightly wider than LSTM, meaning GRU is a bit less accurate for higher cost values.

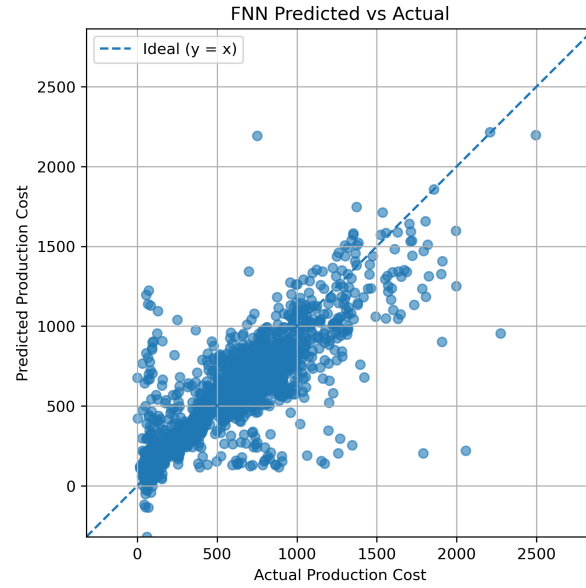


Figure 11: Univariate FNN Predicted vs Actual

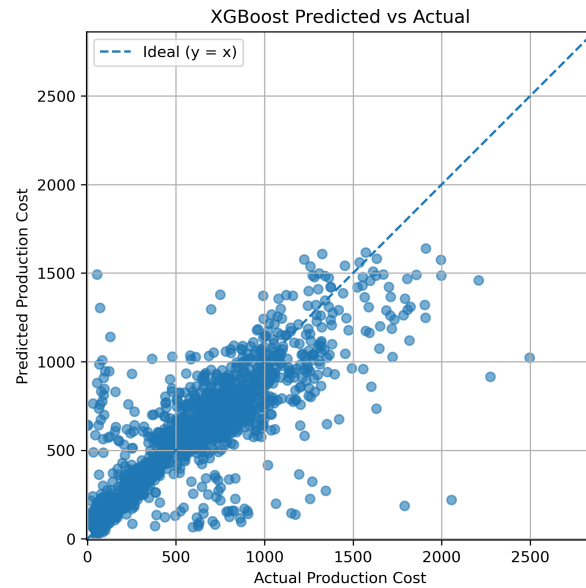


Figure 12: Univariate XGBoost Predicted vs Actual

- The FNN model follows the general pattern but the points are more spread out from the diagonal line. This means the predictions are decent but not very precise, because FNN cannot fully capture day-to-day changes in production cost.
- XGBoost predicts fairly well, and many points are close to the diagonal line. However, there is still some scatter during sudden changes in cost, showing that the model struggles a bit

with sharp spikes or drops.

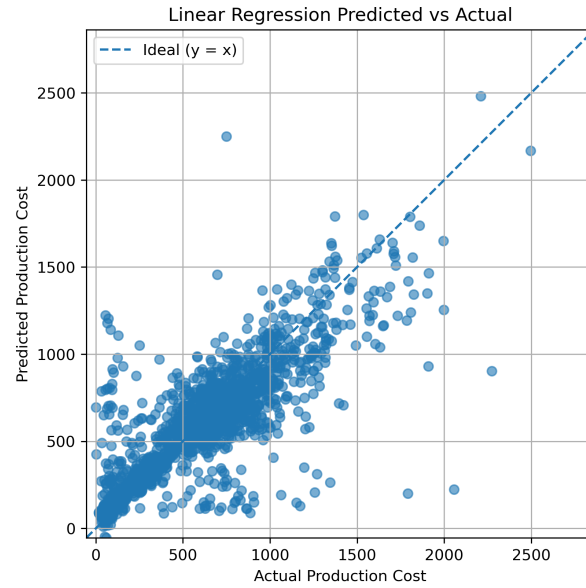
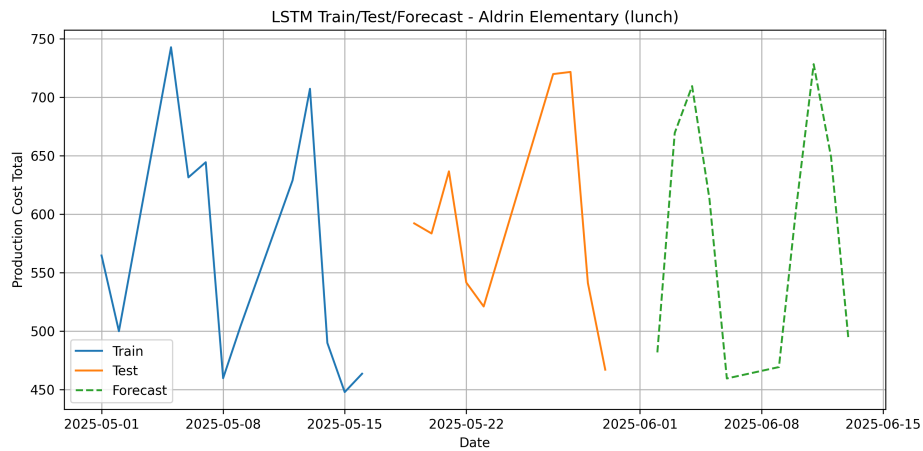


Figure 13: Univariate Linear Regression Predicted vs Actual

- The Linear Regression model captures the general trend, but the points are more spread out from the diagonal line. This means its predictions are less accurate, especially for higher cost values, because a simple linear model cannot learn complex or non-linear patterns in the data.

### 5.3.3 Forecasting Example

Figure 14: Univariate LSTM Train/Test/Forecast Example



The example plot demonstrates:

- Accurate tracking of real cost fluctuations during the test phase

- Stable multi-step forecasts beyond the test horizon
- Smooth adaptation to the underlying temporal trend

This validates the univariate LSTM as the most reliable model for real-world deployment.

## 6 Discussion

This effort aimed at finding out how various machine learning and deep learning models could predict daily school-level production costs using the historical FCPS meal data. In the process, several challenges arose, especially on the complexity and irregularity of production-cost patterns. Unlike datasets used in traditional forecasting studies, for instance, electricity demand or retail product sales, school meal data is highly noisy and inconsistent between schools, as it is often disrupted due to irregular attendance, changes in the menu, and other operational anomalies. These characteristics make it difficult for simpler models to learn stable temporal relationships.

From all models explored, the univariate LSTM was best at overcoming these difficulties. In contrast with related work, which relies on either linear extrapolation or static rule-based approaches, the advantages of using the LSTM were clear because it captured nonlinear dynamics nicely, adapted to cost fluctuations over time, and learned temporal patterns that would not be easily modeled by any classical statistical model. Furthermore, opposed to multivariate models explored in other domains, a univariate approach yielded better results in this project since it avoided noise created by inconsistent operational variables and focused only on clean, reliable cost history.

However, some areas had the alternative solutions outperforming or offering benefits. For example, XGBoost provided stable results with much lower training costs and faster execution times, serving as a very strong baseline for those schools that had limited computational resources. Similarly, GRU networks, even though slightly less accurate compared with LSTM, trained much faster and were not prone to overfitting either; thus, this could be a practical alternative in cases in which quicker model updates were needed.

Despite the robustness of the LSTM, there are limitations to this approach. First, being univariate, it doesn't take into consideration external drivers such as patterns of student attendance, holiday schedules, weather, or even menu-specific cost changes. Second, its performance may be degraded in very long-range predictions; it becomes increasingly sensitive to accumulated error in the model output. These could include integrating external covariates into a multivariate LSTM or transformer-based architecture, seasonal decomposition, and/or accounting for calendar effects that would better adapt to more irregular school periods.

Considering the overall experience from this project, several further approaches can be explored. The attention-based models, such as Transformers and Temporal Fusion Transformers, may better grasp long-range dependencies than the recurrent architecture. Probabilistic forecasting methods include DeepAR or Gaussian Process Regression, which allow quantification of uncertainty and provide confidence intervals for cost estimates. Finally, clustering schools by behavioral patterns before forecasting could reduce noise and improve stability across diverse FCPS schools.

It became evident from the experiments that, although univariate LSTM had the best overall performance in short-term forecasting, the key takeaways from this work also offer the possibility of improving the performance and practical utility of future school meal forecasting systems by incorporating deep learning with contextual features and state-of-the-art sequence models.

## 7 Conclusion

Various machine learning and deep learning models were evaluated in this paper on the forecast of daily production cost in school meal programs. Among the different methods compared, such as Linear Regression, XGBoost, Feedforward Neural Networks, GRU, and LSTM, the univariate LSTM model yielded the best results. It has provided the lowest prediction error and highest stability in all cases, proving a better capability to capture nonlinear and time-dependent patterns of production cost data.

The model developed in this work has yielded accurate short-term forecasts of costs, which would enable more effective budgeting, procurement planning, and reduction of waste in school nutrition programs. The LSTM model effectively learned cost dynamics even in the presence of irregular participation patterns and operational variability, offering a practical solution for daily forecasting needs.

### Overview of Key Results:

- Univariate LSTM provided the highest accuracy with strong  $R^2$  scores with the lowest RMSE values.
- Traditional models captured general patterns but could not model complex temporal behavior.
- The multivariate models did not outperform the univariate LSTM; thus, the univariate approach is the most reliable and efficient.
- The forecasting system successfully generated realistic short-horizon predictions for FCPS schools.

### Remaining Gaps:

- Limited historical data restricts long-term seasonal pattern modeling.
- It does not consider exterior factors like attendance, menu variations, and ingredient pricing.
- Operational disruptions, for example holidays, events, and supply shortages, were not modeled.



## Future Extensions:

- With multivariate features including attendance, menu categories, and pricing trends.
- Exploring longer-horizon forecasting using state-of-the-art architectures like Seq2Seq, Transformers, or N-BEATS.
- Transfer learning to enable improved predictions for schools where less data is available.
- Integrate the forecasting system into real-time school nutrition management platforms for use in everyday operations.

## References

1. Brownlee, J. (2020). *Deep Learning for Time Series Forecasting: Predict the Future with MLPs, CNNs and LSTMs in Python*. Machine Learning Mastery.
2. Chollet, F. (2018). *Deep Learning with Python*. Manning Publications.
3. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780.
4. Yu, H. F., Huang, F. L., & Lin, C. J. (2011). Dual coordinate descent methods for logistic regression and maximum entropy models. *Machine Learning*, 85(1–2), 41–75.
5. Scikit-learn developers. (2024). *Scikit-learn: Machine Learning in Python*. Retrieved from <https://scikit-learn.org/>
6. PyTorch contributors. (2024). *PyTorch Documentation*. Retrieved from <https://pytorch.org/>
7. Kaggle. (2024). *Time Series Forecasting Tutorials*. Retrieved from <https://www.kaggle.com/>
8. Kong, W., Dong, Z. Y., Jia, Y., Hill, D. J., Xu, Y., & Zhang, Y. (2019). Short-term residential load forecasting based on LSTM recurrent neural network. *IEEE Transactions on Smart Grid*, 10(1), 841–851.
9. Lai, G., Chang, W. C., Yang, Y., & Liu, H. (2018). Modeling long- and short-term temporal patterns with deep neural networks. In *Proceedings of the 41st International ACM SIGIR Conference* (pp. 95–104).
10. Mulhollem, J. (2021). U.S. school cafeterias waste more food than those in other developed countries. *Penn State University News*. Retrieved from <https://www.psu.edu/news/>
11. Ralston, K., Treen, K., Coleman-Jensen, A., & Guthrie, J. (2017). *National School Lunch Program: Trends and Factors Affecting Student Participation*. USDA Economic Research Service.
12. Smith, S. L., & Cunningham-Sabo, L. (2014). Food choice, plate waste, and nutrient intake of elementary- and middle-school students in the U.S. National School Lunch Program. *Public Health Nutrition*, 17(6), 1255–1263.