A protocol for meal attendance forecasting at Worcester Academy

Nier Wang[1*]

[1]Worcester Academy, Worcester, U. S. A.

Abstract. In school dining halls, inaccurate meal attendance forecast is the key rationale that leads to either food waste or inefficiencies in which the dining hall runs out of food. The dining service group in Worcester Academy (WA), an independent high school in Massachusetts, United States, has been using some basic forecasting methods required by the contractor company. But it lacks accuracy as the models are not specifically designed for WA, said by the manager. Thus, this study takes a lens into a range of forecasting techniques employed in high schools, universities, and restaurants throughout the world. The techniques include basic time-series models like Naive, Moving Averages, and simple exponential smoothing, to Linear Regression, specialized in ARIMA and Facebook's Prophet, and machine learning algorithms including Lasso regression. The types of data they collected are also identified. Then, the study proposes a new forecasting protocol for WA based on the comparative analysis of methods. This protocol aims for finding the best methods of forecasting at WA. This study will be applied or at least referenced in WA dining service during school years in the future. The best forecasting methods will hopefully be implemented continuously to improve dining hall operations, if feasible.

1 Introduction

Accurate meal forecasting is essential for all types of dining establishments, especially in school settings, as dining hall is the only option for students, faculties, and staff to feed themselves. Excessive food preparation results in waste: according to research, 20% of food produced in hotel industry is wasted under typical circumstances [1]. Conversely, inadequate preparedness leads to hunger of students. Worcester Academy (WA) is an independent high school in Massachusetts. The dining hall at WA simply assumes the meal demand tomorrow would be the same as last time serving the same group of items, as instructed by the contractor dining company SAGE. The dining team only adjusts (prepare more or less) correspondingly based on the special events informed by the school. The forecasts lack accuracy, because they neither take multiple effectual factors into consideration nor are adjusted by recent trends, said the manager. It's meaningful to examine a group of forecasting methods used by other institutions along with their corresponding situations, then choose the ones that can be also implemented at WA based on WA's situation. This study aims to initiate a protocol of implementing meal demand forecasting models at WA in order to have accurate forecasts to

* Corresponding author: nier.wang@worcesteracademy.org

maximize the dining hall operations. This study looks at what the models are, from simple to advanced, used by other organizations, including high schools, universities, and even restaurants, to forecast food demand. This study then emphasizes the methods that perform well in situations (factors in the environment) especially similar to WA as the bases that lead to a clear path developing the protocol.

2 Literature review: comparative analysis of forecasting methods in dining services

2.1 Simple time-series models: Naïve method, moving averages, and simple exponential smoothing

2.1.1 Naïve method

Many institutional food services historically rely on simple forecasting methods. A common naive approach is to assume the next day will be similar to the recent past –tomorrow will be the same as yesterday or the same weekday last week (e.g. if Tuesday, then forecasting same as last Tuesday). While easy to implement, naive methods are often a baseline to improve upon.

2.1.2 Moving averages

Research in a university dining context found that basic mathematical forecasting techniques consistently outperformed the naïve method [1]. In that study, investigators tried forecasting meal counts over 15 weeks using two techniques: moving averages and simple exponential smoothing [1].

Moving averages is a method that calculates the average of attendance numbers from several previous similar days to predict the future turnout (e.g. referencing the past 2–4 Tuesdays to forecast the upcoming Tuesday).

2.1.3 Simple exponential smoothing

Simple exponential smoothing is another technique that updates forecasts by giving weight to recent attendance while still accounting for older patterns. The formula is as follow (F_{t+1} represents the forecast attendance tomorrow):

$$F_{t+1} = \alpha A_t + (1-\alpha)F_t \quad (1)$$

Where,

A_t represents the actual attendance today

F_t represents the forecast for today made yesterday

α is the smoothing constant between 0 and 1 (values subject to how stable the attendance is generally or how responsive the forecasts are expected to be to the fluctuations in attendance recently; the high α, the more responsive).

2.1.4 Comparison

Those simple mathematical forecasting techniques (Moving Averages & Simple Exponential Smoothing) showed its accuracy by the mean absolute percentage error (MAPE) of 6.4% or

less, whereas the naive approach was less accurate [1]. In fact, moving average models gave the best prediction about 83% of the time in the cases studied [1].

In another university (Texas Tech University), moving averages and exponential smoothing were also implemented. And between those two models, moving averages were again superior to more complex exponential smoothing [2]. This sheds light on the idea that sometimes a straightforward model can work better for a given data pattern.

Both methods suggest that using recent historical patterns improves accuracy in university dining halls. Understandably, both models perform well as they both consider the stability of the patterns in the past while constantly adjusting to the new patterns to a reasonable degree. This respects to schools' stable cyclic demand. Since Worcester Academy is a typical high school, both methods can be considered in actual use.

2.2 Linear regression

Upon using all the methods mentioned previously, Texas Tech University (TTU) also tried regression models including simple linear regression and multiple linear regression [2]. A regression model fits the independent variables $x_1, x_2 \ldots x_n$and the dependent variable y with an equation that best represent their relationship in a least-squares sense [3]. Linear regression, either simple or multiple, fits the relationships with a linear line. All linear regression models can be easily implemented by statistical software like R or Python.

2.2.1 Simple linear regression

Simple linear regression models use only one predictor—such as day of the week (used by Ryu & Sanchez in the study of the Texas Tech University cafeteria) — to forecast meal attendance [2]. The form of the equation is:

$$y = \beta_0 + \beta_1 x_1 + \varepsilon \tag{2}$$

Where,

$y \rightarrow$ the predicted attendance

$x_1 \rightarrow$the predictor (a factor that is assumed to be correlated with y)

β_0 , $\beta_1 \rightarrow$coefficients chosen

$\varepsilon \rightarrow$ error term

For example, assume x_1 is the weekday number (Monday=1, Tuesday=2 etc.); plot the 5 participation counts last week, then run linear regression using softwares to find the best fit line; and finally, if one is forecasting for next Tuesday, the y value when x = 2 in the equation is the forecast generated through linear regression.

2.2.2 Multiple linear regression

Multiple linear regression extends this by including multiple predictors instead of one — Ryu & Sanchez use weekday, menu, weather — to forecast meal attendance [2]. If simple linear regression can be visualized by a 2-dimension graph, multiple linear regression is examined in 3-dimension and above. The form of the equation is:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n + \varepsilon \tag{3}$$

Where,

$y \rightarrow$ the predicted attendance

x_1 , x_2 ... x_n →the predictor (a factor that is assumed to be correlated with y)
β_0 , β_1... β_n→coefficients chosen
ε → error term [3].

2.2.3 Comparison

Ryu & Sanchez found that this multiple linear regression model was the most accurate for predicting meal counts [2]. Regression allows the inclusion of multiple factors that are decisive on meal attendance.

The strength of regression-based forecasting especially multiple linear regression is indeed noted by comparison with all other forecasting methods in the study by Texas Tech University due to its considerations of the effects brought by multiple factors, in addition to respecting historical patterns while adjusting as patterns change every day as what simple techniques (moving averages & simple exponential smoothing) do [2]. This indicates that statistical approaches that leverage relevant features (inputs) tend to improve accuracy over one-size fits-all methods [2]. However, notably that Ryu & Sanchez eventually recommended implementing one of the Naïve models, which simply use the attendance from the same weekday of the previous week, to forecast in reality because while multiple linear regression was proven to be the model that has the highest statistical accuracy, NMII has the third highest accuracy while being much easier for dining managers to implement daily [2]. This again emphasizes the idea that while trying to use advanced methods, sometimes a simple naïve model is efficient enough.

Academic schedule, menu offerings and weather conditions should be three of the effectual variables when WA is doing multiple linear regression as they affect meal attendance in all schools. Other factors could be like: yes or no a sports day (when some students miss lunches or dinners), an exam day or week (possibly reducing the number of dinners as students only come to school for days they have exams); for the three factors mentioned above, potential factors are: yes or no a special menu days (Pizza, Chicken tenders days might attract more diners), a snow or rainy day (some students have no appetite for foods under cold weather) etc. Incorporating such factors in the multiple linear regression model may refine the forecasts.

2.3 Advanced time-series models: ARIMA and Facebook's Prophet

In recent years, more advanced forecasting techniques have been tested in foodservice settings to further reduce error and better match supply with demand.

2.3.1 Auto-Regressive Integrated Moving Average (ARIMA)

ARIMA is a classic time-series forecasting model that is used widely. The core idea of ARIMA is that it predicts the changes (Today - the Previous Day) instead of the attendance number directly and eventually adds it to the attendance for the previous day to forecast the number for today. The testing of weights used in the predicted change formula and the following calculations can be done by the help of softwares like R or Python. The formula for the predicted change is as follow:

$$\Delta_t = \varphi_1 \Delta_{t-1} + \varphi_2 \Delta_{t-2} + \ldots + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \varepsilon_t \quad (4)$$

Where
Δ_t →Predicted change

ϕ_1, ϕ_2 → weights for past changes (AR terms)
θ_1, θ_2 →weights for past errors (MA part)
E_t →forecast error
Steps of ARIMA :
(1) Differencing the past values
(2) The computer tests and optimizes the weights, and it then calculates the predicted change using all weights found
(3) Previous day's attendance + predicted change = Today's forecast

2.3.2 Facebook's Prophet

Prophet models the data using the equation

$$y(t) = g(t) + s(t) + h(t) + \varepsilon_t \quad (5)$$

Where,
y(t) → predicted attendance
g(t) →trend
s(t) →seasonality (like weekdays)
h(t) →holiday or special event effects
ε_t →random error
This is also a forecasting method that applies optimization algorithms runned by computers behind the scenes.

ARIMA and Facebook's Prophet are both easy to be implemented using R or Python as long as one has stationary historical data [4]. ARIMA is particularly effective for non-seasonal series, and in a high school setting over a short period of time, the dining attendance does fluctuate around a relatively steady baseline [4]. ARIMA only captures linear relationship as it's a pure time-series model that doesn't account for any other variables like seasonality accounted for in Facebook's Prophet. In a high school setting, the dining hall attendance for a week has seasonality (e.g. there are relatively less people on Wednesdays and Fridays as Wednesdays are the big sports games on which some athletic teams leave for away games early and miss lunch; and for Fridays, some faculties leave earlier if they have no works left), which already causes the data to be non-stationary over a period of time.

2.3.3 Comparison

A 2020 study applied ARIMA and Facebook's Prophet algorithm (along with others) to a university's dining hall data across a year. The study found that these modern time-series methods performed well [5]. Nevertheless, while Prophet was deemed the second-best in accuracy and was praised for its use, notably that the artificial neural network (ANN) model achieved the highest accuracy in predicting meal counts [5]. The ANN's success suggests non-linear patterns and interactions (perhaps between day of week, attendance cycles, etc.) were captured better by a machine learning model in that case.

2.4 Machine learning

Machine learning (ML) methods are increasingly prominent in food demand forecasting. They can handle more complex relationships and larger datasets. Machine learning constitute an alternative approach to classical time-series approach like ARIMA model because while

ARIMA relies on a fixed formula where computer just estimates coefficients to find the best fit, machine learning learns patterns and develop formulas itself between the inputs and the outputs to find the best representation of their relationship [6].

2.4.1 Lasso regression, ensemble methods (XGBoost, Gradient boosting), Linear regression

One very recent study (Turker, 2025) used several ML algorithms on a campus dining dataset, including linear regression, decision trees, Lasso regression, and ensemble methods (XGBoost, Gradient Boosting) [7]. The standout result was that Lasso regression achieved an R^2 of ~0.999, essentially nearly perfect fit for their test data [7]. XGBoost was the next best with R^2 ~0.882 [7]. Ensemble three methods like XGBoost performed better than basic linear regression but still fell short of Lasso's accuracy on that dataset [7]. The likely reason is that Lasso was very effective at picking up the key features (they mentioned features like meal variety, campus population movement, temperature, etc., were relevant [7]. By focusing on the most important variables, Lasso could generalize well [7]. This implies that in that university's case, a well-tuned Lasso model (which is a regularized linear model that selects important features) could predict daily meal counts almost exactly. It's worth noting such a high R^2 for Lasso regression might indicate possible overfitting or an unusually predictable data pattern, but the authors did apply cross-validation to verify generalization. In fact, more complex models risked overfitting noise without careful tuning [7]. ML can dramatically improve forecast precision, especially if one has rich data (multiple features over time). But one must choose the model complexity appropriately in respect to the data volume and variability. Thus, Lasso Regression is overall the best forecasting method to use, but only if the data is rich, including nearly all relevant features even like whether it's a day right after breaks, Day student and boarding student counts, in addition to basic "special data" like menu type and weather.

2.4.2 Random Forest

Another advanced approach applicable to institutional food service is random forest (RF) and related ensemble methods. In a Swedish public school catering study during COVID-19, researchers explored machine learning to forecast meal attendance. They found that a random forest model produced more accurate forecasts than a standard neural network for predicting daily guest counts [8]. The RF model's predictions led to significant economic savings by avoiding overproduction, compared to a scenario with no forecasting or naive forecasting [8]. Although the detailed metrics from that study are complex, it underlines that ML methods can adapt to unusual situations (like the pandemic disruptions) better than static models. Random forests can capture non-linear effects (e.g., the interplay of day-of-week with whether classes were remote or in-person) automatically. The success of such models in a K-12 context (the study included primary school kitchens) is promising for high school use as well.

2.5 Additional notes: forecasting in restaurants

Restaurants of all sizes benefit from accurate forecasting to reduce waste and optimize labor without harming service quality as well [9].

A restaurant industry example highlights the importance of continually updating forecasts and considering external factors--restaurants may adjust by things like local events (concerts, sports games) [9]. In school settings, this could be referred to activity first time happening or rarely happening, like a celebration dinner for Founding Anniversary. This

reminds us of the stakes of forecasting methods. On one hand, they are of great necessity ; on the other hand, we also need to be alert to special situations—different from the "regular" speiciality considered in the models--- that is, significant changes that arise unexpectedly, in which we have to adjust accordingly.

2.6 Summary

2.6.1 Summary of methods of forecasting

In summary, the comparative analysis shows a spectrum of methods:

Firstly, simpler methods (naive, average, smoothing) can be surprisingly effective in steady environments and are easy to implement.

Secondly, statistical models (regression, time-series) bring in more information (trends, factors) and tend to improve accuracy, as demonstrated by multiple studies in university dining settings.

Thirdly, machine learning models (ANN, random forests, boosted trees) offer further gains and can handle complex patterns, though they require more data and expertise.

Lastly, the restaurant sector experience reinforces many of these findings while emphasizing integration of diverse data again and stating the need for continuous refinement of forecasts.

2.6.2 Methods chosen for Worcester Academy

Naïve, moving averages, Simple exponential smoothing: simple, so as baselines,
ARIMA: the most classic method,
Multiple linear regression, Facebook's Prophet, Lasso Regression: they consider multiple factors; additionally, Lasso Regression is theoretically expected to have the highest accuracy.
These lessons will guide the creation of a forecasting protocol suited to Worcester Academy's needs, balancing complexity with practicality.

3 Protocol design for Worcester Academy: finding the best methods of forecasting

As a starting point, the project team should aim the experiment for lunch on WEEKDAYS because everyone eats lunch, while only the boarding community eats breakfast and dinner and the meals on the weekends. After the best methods are figured out, they can be used for the forecasts for the rest of the meals. This protocol covers the outline of gathering data (when to gather what) and implementing forecasting methods (using what and how).

3.1 Data collection plan

Step 1: Create a Google Spreadsheet for data collection daily. List out dates on the first-left column from the top, and factors on the top-first row from the left.

The entire data collection plan takes about 2-3 months, to consider both efficiency and accuracy of the data. Start collecting all of them or as much as possible in the beginning.

3.1.1 Types of data to gather

The only data that Naïve, Moving Averages, Simple Exponential Smoothing and ARIMA need is meal participation counts.

(1) Meal participation counts—how many people eating during lunch

Since Multiple Linear Regression, Facebook's Prophet, and Lasso Regression need data of multiple factors (the more, the better), the project team should potentially consider (as they are suggested in the literature review part)…

(2) Daily Attendance—how many people go to school today

Attendance fluctuates every day and correlates with meal attendance directly.

(3) Menu details-when the popular items are provided

Menu popularity influences turnout. Interview the manager first to have a list of popular items in the school by their serving experience. Record on the spreadsheet when they are served.

(4) Day Type and calendar info—Is today a "regular day" ?

Note whether the day is a regular school day. Record if it is a morning delay, early dismissal day, a sports game day, an assembly day, or if any school-wide special events happening (etc.)

(5) Weather conditions—Does today have "unusual" weather ?

Record the temperature and precipitation for each day. Record potential severe weather conditions like snowstorm or rainstorm. Notably, very cold weather might slightly increase hot meal consumption [7].

(6) Other variables

Pay attention to school life and add any other variables, if found, that are correlated to meal attendance to the spreadsheet. The variables can be the scarcity of a certain type of food in salad bar on a certain day, the first day after school break, or the last day before school break. The spreadsheet should be gradually completed.

3.2 Forecasting tools and methodology

Step 2: Create the second spreadsheet for the record of forecasts (by all methods used), and copy the actual attendance (from spreadsheet 1 for convenience) for each day.

For methods other than naïve and moving averages, R or Python will be used to compute the forecasts.

While the data are gradually collected, here is a potential timeline for starting to implement each forecasting method.

3.2.1 The first 3 weeks (total. W1-12) --- Naïve

The baseline method, as there are not enough data for other methods yet.

3.2.2 The 4th week (t.W4-12) ---Moving Averages, Simple Exponential Smoothing

After three weeks of data collected, enough data (use meal participation counts only here) is there to start moving averages and simple exponential smoothing.

3.2.3 The end of 2nd month (t.W9-12)---ARIMA, Facebook's Prophet

Since we have 8 weeks of data already, ARIMA can now model short-term trends and begin to capture weekly seasonality. Same for the Prophet.

3.2.4 Two weeks after (t.W11-12) ---Multiple Linear Regression, Lasso Regression

Again, machine learning is the best method to use when the data is rich [7]. Thus, we lastly apply Multiple Linear Regression and Lasso Regression.

To avoid overfitting, techniques like cross-validation should be employed – essentially testing the model on unseen data to ensure it generalizes, similar to how researchers validated models across different folds of data [7].

3.3 The end of the project and the future

3.3.1 Evaluating all forecasting methods by MAPE (Mean Absolute Percentage Error)

After 12 weeks of collecting data and implementing forecasting methods, calculate each overall MAPE to measure the accuracy of the forecasts.

$$MAPE = \frac{1}{n}\sum_{t=1}^{n} \left| \frac{A_t - F_t}{A_t} \right| \times 100 \quad (6)$$

Where,

A_t → the actual value in a day

F_t →the forecast for that day (made on the previous day)

n →total observed days over 12 weeks

5-10% MAPE is predicted for moving averages and simple exponential smoothing as that large university cafeteria has 6.4% [1].

3.3.2 Ranking all methods by accuracy and complexity

Recall that Ryu & Sanchez in Texas Tech University eventually recommended one of the Naïve methods since NMII (the name of this method) has the third highest accuracy while being much easier for dining managers to implement daily [2].

After calculating MAPE, rank all methods by MAPE from the highest to the lowest. Meanwhile, rank the complexity of implementing as well through experience (how hard it was to collect all data needed and using R or Python to run the program).

Keep in mind that simple methods might perform better, and even not so, the accuracy is possibly only slightly off while much easier to implement.

3.3.3 Choosing methods for the future of WA

After ranking, the methods that should be continuously used in the future will be obvious.

3.3.4 The future

The project club can consider a simple approach for long-term implementation of the forecasting methods. Possibly, the team could develop a user-friendly program for the dining hall staff to use, with the core functionality based on the selected forecasting models, as it will be difficult and confusing for the dining team to use R or Python every day.

In effect, the dining hall will operate more like a well-tuned restaurant or university cafeteria, using evidence-based forecasts to guide purchasing and cooking – "ensuring the right products are available at the right time" while minimizing excess [10].

4 Conclusion

Naïve, Moving Averages, Simple Exponential Smoothing, ARIMA, Multiple linear regression, Facebook's Prophet, and Lasso Regression are the forecasting methods suggested

for Worcester Academy's dining services after examining how other institutions including high school, universities, and even commercial restaurants forecast their meal attendance. They are chosen due to the fact that they are conventional, typical, accurate, and generally appropriate. A protocol which touches upon the types of data to gather and the timeline of using the forecasting tools is developed afterwards. The best forecasting method(s) will be chosen for future use after ranking them all by accuracy (MAPE) and complexity(experience). Once the validated methods are implemented continuously after the project, a more productive dining team will produce higher-quality food by maximizing budget usage. In terms of long-term use, creating a user-friendly computer program for the dining team is the next goal. As a result, satisfaction with the dining experience will increase, which may positively impact everyone's performance at school, yielding significant progress at WA. Along with that, implementing this project will foster a culture of data-informed decision making as well. These factors might be useful in adjusting the forecasts correspondingly if observations impli that they will bring the project team and the school closer to the goals.

References

1. L. Blecher, R. J. Yeh, Forecasting Meal Participation in University Residential Dining Facilities. Journal of Foodservice Business Research. **11(4)**, findings on moving average vs. naive (2008) forecastsresearchgate.net
2. K. Ryu, A. Sanchez, The Evaluation of Forecasting Methods at an Institutional Foodservice Dining Facility – reported multiple linear regression as the best (2003) modelresearchgate.net
3. R. Bevans, Multiple Linear Regression | A Quick Guide (Examples), Scribbr (2023) www.scribbr.com/statistics/multiple-linear-regression/
4. K. Raj, Ultimate Guide for ARIMA Time Series Forecasting, Dataaspirant (2023) https://dataaspirant.com/arima-time-series-forecasting
5. M. Yurtsever, V. Tecim, Forecasting Meal Requirements Using Time Series Methods in Organization – showed ANN outperforming ARIMA and Prophet for university dining (2020) dataideas.repec.org
6. A. Kontopoulou, et al. A Review of ARIMA vs. Machine Learning Approaches for Time Series Forecasting in Data Driven Networks. Future Internet. **vol. 15**, no. 8, 255 MDPI (2023) https://doi.org/10.3390/fi15080255
7. G. F. Turker, Reducing Food Waste in Campus Dining: A Data-Driven Approach to Demand Prediction and Sustainability – Sustainability, **17(2)**, 379. Demonstrated high accuracy of Lasso regression (R^2 ~0.999) and significant food waste reduction with ML (2025) modelsmdpi.commdpi.com
8. C. Malefors, et al., Food waste reduction and economic savings… machine learning in Swedish public catering – Socio-Economic Planning Sciences, **82**, highlighted random forest forecasts improving accuracy and saving costs in school meal (2022) programsresearchgate.netresearchgate.net
9. NetSuite Business Guide, Restaurant Forecasting: How-To with Formulas – Emphasizes the role of data and technology in restaurant sales forecasting and its benefits for inventory and labor (2025) managementnetsuite.comnetsuite.com
10. Tastewise, How Food Demand Forecasting is Shaping the Future of Foodservice – Notes that modern analytics (AI) and external factors are increasingly used to refine food demand (2024) predictionstastewise.iotastewise.io