

## Development of Machine Learning-Based Predictive Models and a Web App for University Cafeteria Attendance

Gwangwon Jung\*, Keewon Kim\*\*

\*AI Research Engineer, StudioM Co., Ltd, Suwon, Korea

\*\*Professor, Dept. of Computer Engineering, Mokpo National Maritime University, Mokpo, Korea

### [Abstract]

Accurate meal attendance prediction positively impacts cafeteria operations by reducing food waste and lowering costs. In the past, predictions were based on human experience, but modern methods involve developing predictive models. These models offer advantages such as reduced time and resource consumption compared to traditional methods. However, practical implementation remains limited due to various challenges. This paper utilizes cafeteria data from 2018, 2019, and 2023 semesters, along with weather data from the Korea Meteorological Administration, to select variables through EDA analysis. Using these variables and machine learning algorithms, models were developed to predict breakfast, lunch, and dinner attendance. The best-performing model was selected as the final model to develop a prediction program. 80% of the data was used for training and 20% for validation. The final models selected were XGBoost for breakfast (MAE ratio 12.97%) and RandomForest for lunch and dinner (MAE ratios 4.8% and 4.93%, respectively). The lunch and dinner prediction models demonstrated good performance. Future work will involve continuous data collection, adding new derived variables, hyperparameter tuning, and UI improvements to enhance model performance and convenience.

▶ **Key words:** Cafeteria, Predictive Model, EDA, Machine Learning, Deep Learning

### [요약]

정확한 식사 인원 예측은 잔반 감소와 원가 절감 등 식당 운영에 긍정적인 영향을 미친다. 과거에는 사람의 경험에 의존하여 예측했으나, 현대에는 예측 모델을 개발하여 사용하고 있다. 이러한 예측 방법은 과거의 방식보다 적은 시간과 자원을 소모하는 등의 이점이 있다. 그러나 여러 어려움으로 인해 실제로 이를 시도하는 곳은 많지 않다. 이전 연구에서의 모델을 고도화하여 2018년, 2019년, 2023년 학기의 교내 식당 데이터와 기상청의 날씨 데이터를 사용하여 EDA 분석을 통해 변수를 선정하였다. 머신러닝 알고리즘을 사용하여 조식, 중식, 석식의 식사 인원을 예측하는 모델을 개발하였고, 전체 데이터의 90%로 학습하고 나머지 10%를 검증에 사용하였다. 가장 성능이 우수한 모델을 최종 모델로 선정하여 예측 프로그램을 개발하였다. 최종 모델은 조식 예측에 XGBoost(MAE 비율 12.97%), 중식과 석식 예측에 RandomForest(MAE 비율 각각 4.8%, 4.93%)를 사용하였다. 중식과 석식 예측 모델이 좋은 성능을 보였다. 향후 지속적인 데이터 수집, 새로운 파생 변수 추가, 하이퍼파라미터 튜닝, UI 개선을 통해 성능과 편의성을 더욱 향상시킬 예정이다.

▶ **주제어:** 구내식당, 예측 모델, 탐색적 데이터 분석, 기계 학습, 딥러닝

• First Author: Gwangwon Jung, Corresponding Author: Keewon Kim

\*Gwangwon Jung (rhd5148@gmail.com), StudioM Co., Ltd

\*\*Keewon Kim (kwkim@mmu.ac.kr), Dept. of Computer Engineering, Mokpo National Maritime University

• Received: 2024. 07. 09, Revised: 2024. 11. 14, Accepted: 2024. 11. 14.

## I. Introduction

특정 대상들에게 주기적으로 음식을 제공하는 식당에서 식사 인원 예측은 식재료 낭비 감소, 서비스 품질 향상, 다양한 메뉴 제공, 그리고 효율적인 재고 관리를 가능하게 한다. 따라서 식사 인원 예측을 통해 식당 운영의 효율성과 서비스 품질이 향상될 수 있다[1].

과거에는 구내 식당의 식수 인원을 예측하기 위해 사람이 여러 통계 공식을 직접 계산하거나 통계 프로그램을 통해 계산하였다. 이러한 방식은 대용량 데이터를 처리할 때 많은 시간과 자원을 소모하였다. 또한 예측 과정에서 사람의 주관이 개입되어 오류가 발생하거나 일정한 패턴과 규칙을 발견하기 어려운 경우가 많았다. 이러한 문제를 해결하기 위해 최근에는 인공지능 기술을 활용한 식수 인원 예측 모델이 개발되고 있다[2]. 이 모델들을 식당 운영에 적용하면 잔반 감소, 원가 절감, 음식 품질 향상 등 다양한 긍정적인 효과를 얻을 수 있다.

실제 대기업이나 다른 업체 또는 기관의 구내식당을 주주하여 운영하는 곳에서는 식당 운영 데이터를 활용해 통계 분석 및 예측 모델을 사용하고 있다[3-6]. 그러나 대부분의 식당은 여전히 식당 운영자의 경험에 의존하여 식사 인원을 예측하고 있다. 이러한 현상의 원인은 다음과 같다. 실제로 사용할 수 있는 식사 인원 예측 모델을 개발하기 위해서는 각 식당에 맞는 데이터가 충분히 수집되어야 하며, 그 데이터를 기반으로 유의미한 성능을 보이는 모델을 개발할 수 있는 전문가가 필요하다. 또한, 식당마다 식사 인원수에 영향을 미치는 요인이 다르기 때문에 각 식당에 맞는 새로운 모델을 개발해야 하는 어려움이 있다[7].

이전 연구 [8]에서는 2018년과 2019년 학기 중 식당 데이터와 기상청 데이터로 XGBoost 모델을 학습하여 식사 인원 예측 모델을 생성하였다.

본 논문에서는 이전 연구에서 비해, 예측 모델 학습에 2023년의 모든 학기 데이터를 추가하였고, 데이터전처리를 통해 추가적인 변수를 생성하고 불필요한 데이터를 삭제하여 더 정제된 데이터를 생성하였다. 그리고 Random Forest 알고리즘과 DNN과 LSTM 같은 딥러닝 알고리즘들을 비교 알고리즘에 추가하였다. 2018년, 2019년, 2023년 학기 중 교내식당에서 얻은 식당 데이터와 기상청의 날씨 데이터를 EDA 분석 및 전처리 과정을 거쳐 Machine Learning과 Deep Learning 알고리즘을 사용하여 조식, 중식, 석식의 식사 인원을 예측하는 모델을 개발한다. 이후, 개발된 모델을 바탕으로 실제 학교 식당에서 영양사가 사용할 수 있는 프로그램을 제작한다. 이 프로그램을 통해

교내식당에서 발생하는 음식 품질과 잔반 문제를 개선하여 식당 운영에 긍정적인 효과를 주고자 한다.

본 연구의 구성은 다음과 같다. 제2장에서는 식사 인원을 예측하는 모델의 개발에 대해 상세히 설명한다. 제3장에서는 이 모델을 활용하여 식사 인원을 예측하는 프로그램에 대해 소개한다. 제4장에서는 이 예측 모델의 성능을 평가하고 그 결과를 제시한다. 마지막으로, 제5장에서는 이 연구의 결론과 함께 향후 연구 방향을 제안한다.

## II. Develop Models for Estimating Diner Number in a University Cafeteria

### 1. Data and Data Preprocessing

본 연구에서 사용하는 학습, 검증, 테스트 데이터는 2018년, 2019년, 2023년 학기 중 해당하는 날짜의 실제 학교 식당 데이터와 기상청의 날씨 데이터이다. 식당 데이터는 국립목포해양대학교 구내식당의 영양사로부터 제공 받은 것으로, 2018년 229개, 2019년 213개, 2023년 216개로 총 658개의 데이터가 있다. 식당 데이터의 구성은 Table 1과 같이, '날짜', '요일', '조식 메뉴', '조식 인원', '중식 메뉴', '중식 인원', '석식 메뉴', '석식 인원', '총원', '행사'의 10개 컬럼으로 구성된 시계열 데이터이다.

Table 1. University Cafeteria Data

| no | Column               | Type     |
|----|----------------------|----------|
| 1  | Date                 | Datetime |
| 2  | Day of the week      | String   |
| 3  | Breakfast menu       | String   |
| 4  | Breakfast attendance | Integer  |
| 5  | Lunch menu           | String   |
| 6  | Lunch attendance     | Integer  |
| 7  | Dinner menu          | String   |
| 8  | Dinner attendance    | Integer  |
| 9  | Total attendance     | Integer  |
| 10 | Events               | String   |

날씨 데이터는 기상자료개방포털[9]에서 2018년, 2019년, 2023년 학기 중 해당하는 날짜의 목포 지역의 기온, 강수량, 습도 데이터를 이용한다. 각 데이터들은 2018년, 2019년, 2023년 전체 일수를 포함하기 때문에 총 1095개가 있다. 날씨 데이터의 구성은 Table 2와 같이 온도, 강수량, 습도로 구성된다.

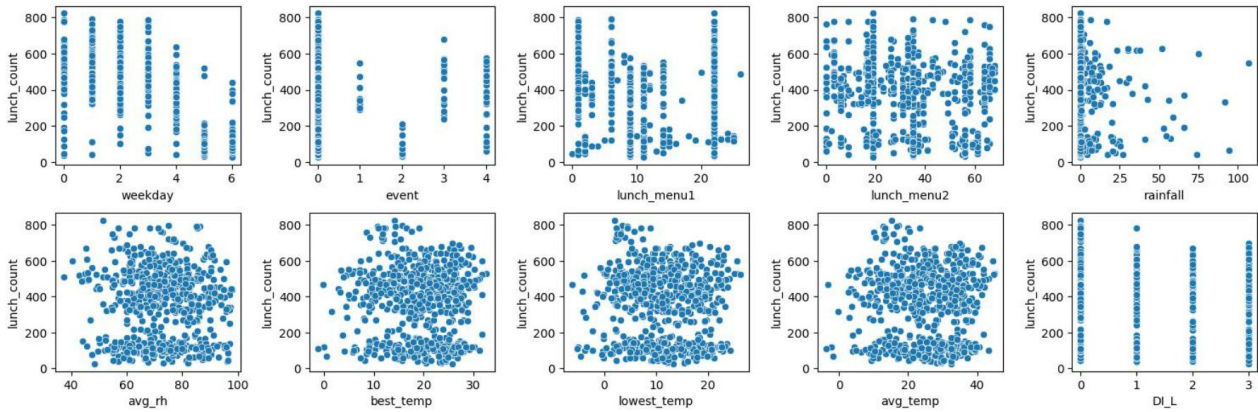


Fig. 1. Visualization of the number of lunch diners for a column

Table 2. Weather data

| Category    | No. | Column                          | Type     |
|-------------|-----|---------------------------------|----------|
| Temperature | 1   | Date                            | Datetime |
|             | 2   | Average temperature             | Float    |
|             | 3   | Maximum temperature             | Float    |
|             | 4   | Time of the maximum temperature | Time     |
|             | 5   | Minimum temperature             | Float    |
|             | 6   | Time of the minimum temperature | Time     |
|             | 7   | Diurnal variation               | Float    |
| Rainfall    | 1   | Date                            | Datetime |
|             | 2   | Precipitation                   | float    |
| Humidity    | 1   | Date                            | Datetime |
|             | 2   | Average humidity                | float    |
|             | 3   | Minimum humidity                | float    |

메뉴 데이터는 콤마를 기준으로 여러개의 메뉴가 하나의 문자열로 이루어져 있고 “밥”, “국”, “메인 메뉴”, “서브 메뉴1”, “서브 메뉴2”, ... , “부식” 형태로 구성된다. 학생들의 메뉴 선호도 조사를 했을 때, 부식이 가장 선호도가 높았고 그 다음 메인 메뉴, 서브 메뉴, 밥, 국 순이었다. 이 중 부식과 메인 메뉴를 선호 메뉴1, 선호 메뉴2로 정하였다.

기온, 강수량, 습도 데이터를 일시를 기준으로 병합하고, 이후 '날짜', '강수량', '평균 상대습도', '최고기온', '최저기온', '평균기온'을 사용하였다.

평균 상대습도, 최고기온, 최고기온 시각, 최저기온, 최저기온 시각, 평균기온을 기반으로 불쾌지수(discomfort index)를 계산하고, 이를 기상청에서 정의한 불쾌지수 단계 [10]에 따라 분류하였다. 본 논문에서 사용하는 불쾌지수를 구하는 식은 Thom [11]이 제안한 식 (1)과 같다.

$$DI = \frac{9}{5}T - 0.55(1 - RH)\left(\frac{9}{5}T - 26\right) + 32 \quad (1)$$

여기서  $DI$ 는 불쾌지수이고,  $T$ 는 기온(°C)이고,  $RH$ 는 상대습도(%)이다.

전처리된 조식, 중식, 석식의 식당 데이터를 전처리된 날씨 데이터와 날짜를 축으로 병합하였다. 최종적으로 데이터를 훈련과 검증을 위해서 병합된 데이터에서 각각 90%와 10% 비율로 학습 세트와 검증 세트로 분할하였다.

## 2. Variable Selection

본 연구에서는 변수 선택을 위해서 전처리된 데이터셋을 활용하여 탐색적 데이터 분석(EDA)을 수행하였다. 조식, 중식, 석식의 식수 인원 변화에 영향을 미치는 주요 요인들을 시각화 기법을 통해 파악하였다.

EDA 분석 결과를 요약하면 다음과 같다. Fig. 1은 각 컬럼에 대한 점심 식사 인원수를 시각화한 것이다. 요일에 따른 식사 인원수의 시각화를 살펴보면 가로축의 0~6은 월요일부터 일요일까지 순서대로 매핑한 것이다. 평일과 주말은 식사 인원수의 차이가 명확하게 보이고, 주말로 갈수록 식사 인원수가 줄어드는 것을 볼 수 있다.

행사에 따른 식사 인원수의 시각화를 살펴보면, 가로축의 0~4은 각각 “없음”, “견학”, “휴일”, “중간고사”, “기말고사”를 매핑 한 것이다. “없음”일 때는 식사 인원수가 고르게 분포되어 있으나, 다른 행사의 경우 중간 또는 적은 인원수에 분포된 것을 볼 수 있다.

메뉴에 따른 식사 인원수의 시각화를 살펴보면, 가로축의 값들은 메뉴1에서는 부식, 메뉴2에서는 메인 메뉴를 숫자로 매핑 한 것이다. 메뉴1을 보면 특정 메뉴에 따라 인원이 적어지거나 많아지는 경우가 있다. 메뉴2도 메뉴1과 동일한 패턴으로 특정 메뉴에 따라 인원이 적거나 많아지는 경우가 있다.

강수량에 따른 식사 인원수의 시각화를 살펴보면, 가로축의 값들은 강수량이다. 강수가 없을 때는 고르게 인원이

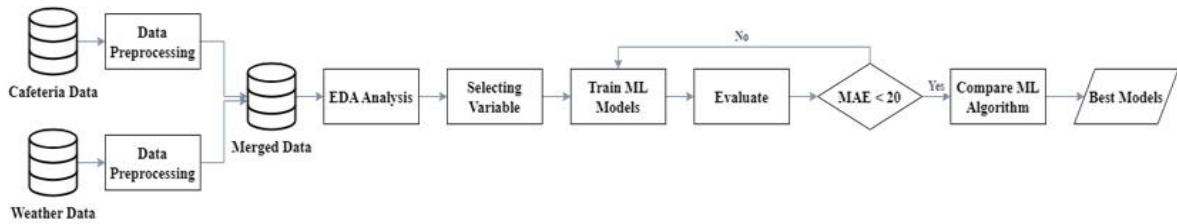


Fig. 2. Models Development Process

있지만 강수량이 증가할수록 상대적으로 적은 식사 인원인 것을 알 수 있다.

평균 상대습도에 따른 식사 인원수의 시각화를 살펴보면, 가로축의 값들은 평균 습도이다. 평균 습도가 60에서 90 정도의 구간에서 식사 인원이 고르게 분포 되었고 나머지 구간에서는 많거나 적은 식사 인원이 적은 수로 분포되어 있다.

최고 온도, 최저 온도, 평균 온도에 따른 식사 인원수의 시각화를 살펴보면 가로축의 값들은 각각 최고 온도, 최저 온도, 평균 온도이다. 이들은 비슷한 형태의 분포를 가진다. 온도의 중간에서 조금 왼쪽의 범위에서 제일 많은 식사 인원수들이 몰려있다. 200명에서 300명 사이의 식사 인원수는 상대적으로 적다. 위 두 가지 특징을 제외한 나머지 범위는 고르게 분포되어 있다.

불쾌지수에 따른 식사 인원수의 시각화를 살펴보면 가로축의 값들은 각각 “low”, “normal”, “high”, “vhigh”를 숫자에 매핑한 것이다. 불쾌지수가 low일 때는 식사 인원수가 적은 수부터 제일 많은 수까지 고르게 분포되어 있다. 그러나 나머지 “normal”, “high”, “vhigh”의 경우는 low에 비해 제일 많은 수가 상대적으로 적다.

EDA 분석을 통해 각 데이터의 식사 인원수에 미치는 요인, 데이터의 분포 패턴 등을 종합적으로 고려해 “요일”, “행사”, “선호 메뉴1”, “선호 메뉴2”, “강수량”, “최고 기온”, “최저 기온”, “평균 기온”, “평균 상대습도”, “불쾌지수 단계”를 독립 변수로 선정하였다.

**3. Models Development**

본 연구에서 제시하는 모델 개발 과정은 Fig. 2에 나타나 있다. 우선, 식당 데이터와 날씨 데이터를 수집한다. 이 데이터는 전처리 과정을 통해 데이터 정제 과정을 수행한다. 다음으로 정제된 데이터를 통합하여 조식, 중식, 석식 각각에 대한 통합 데이터를 생성한다. 이렇게 생성된 데이터들은 탐색적 데이터 분석(EDA)을 통해 데이터의 특성을 파악하고, 이를 바탕으로 새로운 파생 변수를 생성한다. 이후 모델 학습에 사용할 변수를 선정하고, 비수치 데이터

의 경우 다음과 같이 수치형 데이터로 변환한다. 메뉴1과 메뉴2의 경우 메뉴 이름 사전을 만들어 매핑하였다. 요일, 행사, 불쾌지수의 경우는 종류가 많지 않으므로 순서대로 숫자 0부터 1씩 증가하도록 매핑하였다. 마지막으로, 변환된 데이터를 훈련과 검증 세트로 분리한다.

Table 3. Hyper Parameters

| Model                    | Hyperparameters     | Value  |
|--------------------------|---------------------|--------|
| Random Forest Regression | n_estimators        | 7000   |
|                          | max_depth           | 10     |
|                          | min_samples_split   | 2      |
|                          | min_samples_leaf    | 1      |
|                          | max_features        | “auto” |
| XGBoost Regression       | n_estimators        | 8000   |
|                          | learning_rate       | 0.05   |
|                          | max_depth           | 6      |
| DNN                      | min_child_weight    | 1      |
|                          | Activation Function | “Adam” |
|                          | Batch Size          | 32     |
|                          | Learning Rate       | 0.003  |
| LSTM                     | Epoch               | 10000  |
|                          | Batch Size          | 32     |
|                          | Learning Rate       | 0.002  |
|                          | Sequence Length     | 50     |
|                          | Epoch               | 10000  |

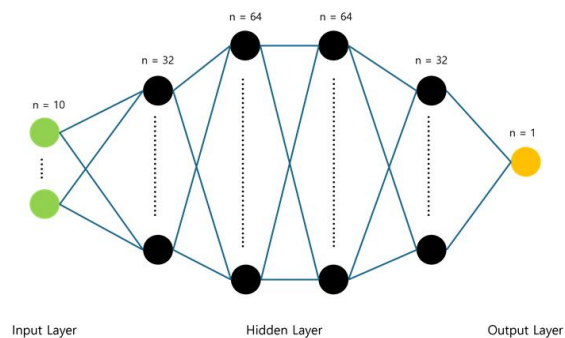


Fig. 3. Our DNN Model Architecture

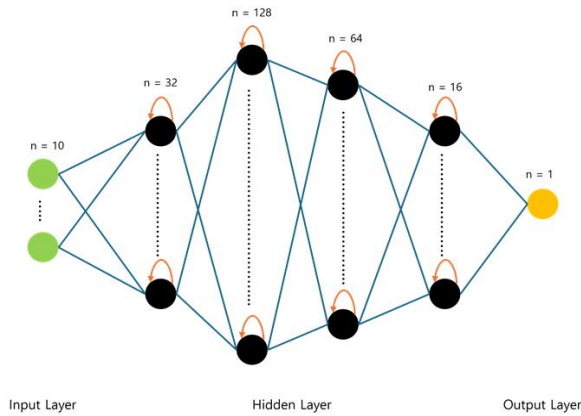


Fig. 4. Our LSTM Model Architecture

본 연구에서는 RandomForest Regression, XGBoost Regression, DNN, LSTM의 알고리즘을 활용하여 모델을 생성하고 학습 데이터셋을 사용하여 학습을 진행한다.

RandomForest Regression은 여러 개의 Decision Tree를 앙상블하여 예측을 수행하는 알고리즘이다. XGBoost Regression은 Boosting 알고리즘의 하나로, 여러 개의 약한 학습기를 앙상블하여 강력한 예측 모델을 만드는 방법이다. DNN은 여러 개의 은닉층을 가진 인공신경망 구조로 다양한 데이터에 복잡한 패턴을 학습할 수 있다. LSTM은 RNN의 일종으로, 시계열, 문장 등 순서가 있는 데이터를 처리하는 데 특화된 알고리즘이다.

RandomForest Regression과 XGBoost Regression은 Random Search를 사용하여 하이퍼파라미터 튜닝을 진행하여 최적의 하이퍼파라미터를 구한다. 각 하이퍼파라미터의 임의의 값 조합 중에서 Score가 가장 좋은 조합을 선택하였다. 각 모델의 최적 파라미터 값은 Table 3과 같다. DNN과 LSTM의 구조는 Fig. 3과 Fig. 4와 같다.

모델의 성능 평가는 평균 절대 오차(MAE: Mean Absolute Error)를 사용한다. 검증 세트를 사용하여 학습된 모델의 성능을 평가하며, 이때 MAE 비율값이 20 미만인 경우, 하이퍼파라미터를 조정하여 모델을 다시 학습한다. 만약 MAE 비율값이 20 미만이라면, MAE 비율값, RMSE, R-Squared 평가지표를 구한다.

#### 4. Evaluation Methods for Models

식사 인원 예측은 실제값의 규모에 따라 모델 성능을 공정하게 비교하기 어렵기 때문에 모델 평가지표로 MAE의 비율을 사용한다. MAE는 식 (2)와 같이 계산된다[12].

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (2)$$

여기서  $y_i$ 는 실제값,  $\hat{y}_i$ 는 모델 예측값,  $n$ 은 데이터 개수이다.

MAE의 비율은 MAE 값을 실제값의 평균으로 나눈 후 100을 곱하여 퍼센트로 표현한 지표이다. 모델 학습은 이 비율이 20% 이하가 될 때까지 하이퍼파라미터를 튜닝하며 반복적으로 수행하였다.

### III. Development of a System Estimating Diner Number

#### 1. Training Prediction Models

모델 훈련은 이전에 설명한 네 가지 알고리즘을 사용하여 진행되었다. 훈련을 시작하기 전에 수치형이 아닌 데이터를 처리하고 데이터 스케일링을 수행했다. 데이터 스케일링은 다음과 같이 진행되었다.

독립 변수는 “요일”, “행사”, “선호 메뉴1”, “선호 메뉴2”, “강수량”, “불쾌지수”, “평균 상대습도”, “최고 기온”, “최저 기온”, “평균 기온”로 이루어져 있다. 독립 변수 데이터는 먼저 MinMaxScaler를 사용하여 값을 0과 1 사이로 조정했다. 그런 다음 IQR을 활용하여 MIN(20%)보다 작거나 MAX(20%)보다 큰 값을 이상치로 간주하고, MIN보다 작은 값은 MIN 값으로 대체하고, MAX보다 큰 값은 MAX 값으로 대체했다. 마지막으로 기온과 관련된 데이터는 다분포를 보여 정규분포와 유사하게 변환하기 위해 데이터 평균을 빼고 절대값을 취했다.

종속 변수는 조식, 중식, 석식 각각 시간대의 식사 인원이다. 종속 변수 데이터는 오른쪽으로 치우친 형태를 보이기 때문에 전체 데이터에 제곱근 및 로그 변환을 적용하여 양의 왜도를 감소시켜 대칭 분포로 변환했다. 또한 MinMaxScaling과 IQR을 통해 이상치를 제거했다.

이렇게 스케일링한 데이터를 사용하여 조식, 중식, 석식에 각각 네 가지 알고리즘 모델을 학습하였다.



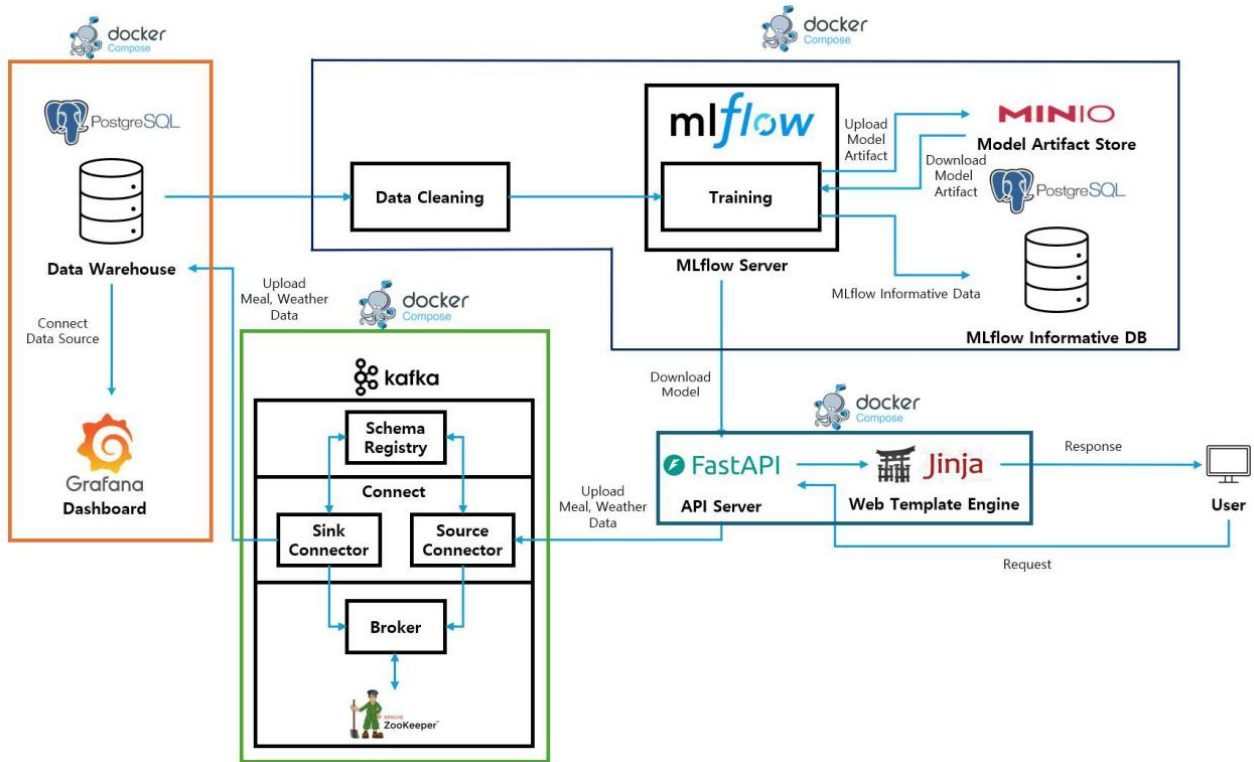


Fig. 5. System Architecture

Table 4. Program Development Environment

| Item        | Value   |
|-------------|---|
| OS          | MacOS(Sonoma 14.2.1)  |
| IDE         | Visual Studio Code for Mac                                  |
| Language    | Python 3.9.0  |
| Library     | Tensorflow, Sklearn, MLflow, Pandas, Numpy, FastAPI, Jinja2 |
| Storage     | PostgreSQL, MinIO   |
| Dashboard   | Grafana   |
| Running Env | Docker Container  |

## 2. System Development

본 연구의 시스템 개발 환경은 Table 4와 같다. 본 연구에서는 식수 인원을 예측하는 웹 앱을 개발했다. 개발한 시스템의 전체 구성은 Fig. 5와 같다. 각 서버는 컨테이너로 구성되어 있으며 동일한 Docker 네트워크 내에서 통신한다.

모델의 학습과 관리를 자동화하고 효율적으로 관리하기 위해 MLflow, MinIO, 그리고 PostgreSQL을 사용하여 MLflow Model Registry를 구성했다. MLflow를 통해 학습 과정을 추적하고 Model Life Cycle을 관리한다[15].

API 개발에는 FastAPI를 활용했다. 식수예측 모델의 학습과 예측을 Python을 사용하므로 호환성이 좋은 Python Web Framework 중 AI에 특화 서비스에 특화되어 있는 FastAPI를 선택했다[16].

사용자가 예측할 식사 시간대와 관련 데이터(날짜, 행사, 주요 메뉴1, 주요 메뉴2, 강수량, 평균 상대습도, 최고 기온, 최저 기온, 평균 기온)가 저장된 .xlsx 파일을 서버로 전송하면, 서버에서 데이터를 모델 예측에 사용할 수 있도록 전처리한다. 선택한 식사 시간대에 맞는 모델에 데이터를 입력하면 FastAPI 서버는 MLflow를 통해 해당 모델을 가져온다. 전처리 프로세스를 통해 입력 데이터를 예측에 적합한 형태로 변환한 후 모델을 활용하여 예측을 수행한다. 최종 예측 결과는 엑셀 파일로 바탕화면에 자동으로 저장된다.

학습 데이터 전달에는 Kafka [17]를 활용했다. Kafka를 사용하면 데이터를 실시간으로 처리하고 원하는 위치로 효율적으로 라우팅하여 전달할 수 있다. 사용자가 학습 데이터를 추가하면 FastAPI 서버는 해당 데이터를 받아 Kafka 시스템을 통해 Data Warehouse에 추가한다.

데이터 시각화는 Grafana [18]를 활용했다. Grafana는 오픈 소스로 다양한 데이터베이스를 연결하여 대시보드 형태로 데이터를 시각화할 수 있다. 사용자는 Data Warehouse에 저장된 데이터를 Grafana 대시보드를 통해 시각적으로 확인할 수 있다.



Fig. 6. Lunch Predict Page

### 3. Result of the Developed Web App

사용자는 각 데이터에 해당하는 엑셀 파일을 업로드하여 데이터를 추가할 수 있다. 이렇게 추가된 데이터는 Kafka를 통해 Data Warehouse로 전달되어 저장된다.

Fig. 6은 점심 예측 페이지를 보여준다. 사용자는 "조식, 중식, 석식" 중 예측하고자 하는 값을 선택할 수 있으며, 선택한 값에 따라 학습된 모델이 자동으로 선택된다. 입력 변수로는 "날짜, 행사, 주요 메뉴1, 주요 메뉴2, 강수량, 평균 상대습도, 평균 기온"을 사용하여 단일 값 예측 또는 엑셀 파일을 업로드하여 예측할 수 있다. 예측 결과는 예측 날짜와 예측 인원이다.

| 날짜         | 실제값  | 예측값  | 오차의 절댓값 |
|------------|------|------|---------|
| 2023.04.28 | 46명  | 35명  | 11명     |
| 2023.04.30 | 420명 | 425명 | 5명      |
| 2023.05.14 | 380명 | 382명 | 2명      |
| 2023.05.16 | 455명 | 413명 | 42명     |
| 2023.05.20 | 25명  | 27명  | 2명      |
| 2023.05.24 | 420명 | 427명 | 7명      |
| 2023.05.30 | 360명 | 311명 | 49명     |

Fig. 7. The Result of Lunch Diners Prediction Comparison Page

Fig. 7은 중식 예측결과를 나타낸다. 결과 예측 비교 페이지는 MAE 확률과 날짜, 실제값, 예측값, 오차의 절댓값을 보여준다.

## IV. Result of Models Evaluation and Program Testing

모델 평가는 검증 데이터로 진행한다. 각 모델별로 MAE 비율, RMSE, R-Squared 평가 지표를 구하였다. Table 5는 모델별 평가 결과를 나타낸 표이다.

Table 5. Result Model's Evaluations

| Meals      | Algorithm     | MAE    | RMSE | R-Squared |
|------------|---------------|--------|------|-----------|
| Break fast | Random Forest | 13.58% | 17명  | 0.60166   |
|            | XGBoost       | 12.97% | 18명  | 0.67312   |
|            | DNN           | 14.63% | 19명  | 0.5433    |
|            | LSTM          | 19.7%  | 26명  | 0.49177   |
| Lunch      | Random Forest | 4.8%   | 92명  | 0.99982   |
|            | XGBoost       | 5.17%  | 92명  | 0.9996    |
|            | DNN           | 12%    | 215명 | 0.662     |
|            | LSTM          | 11%    | 197명 | 0.6144    |
| Dinner     | Random Forest | 4.93%  | 51명  | 0.99971   |
|            | XGBoost       | 5.21%  | 51명  | 0.95237   |
|            | DNN           | 11.05% | 113명 | 0.7022    |
|            | LSTM          | 7.4%   | 74명  | 0.9046    |

먼저 MAE 비율을 살펴보면 실제값 대비 오차가 차지하는 비율이 어느 정도인지 알 수 있다. RMSE는 오차가 몇 명인지 직관적으로 알 수 있지만 실제 데이터에서 각 시간대 식사 인원이 일반적으로 몇 명인지 알아야 모델의 성능을 평가할 수 있다. R-Squared는 값이 1에 가까울 수록 모델이 좋은 성능을 보이기 때문에 직관적으로 알 수 있지만 과적합에 약하고 비선형 데이터에 약한 평가지표이다. 이러한 특징들을 종합적으로 고려했을 때 모델을 평가하는 지표로 MAE 비율을 선정하였다.

모델 평가 결과, 조식은 MAE 비율값이 12.97%, 중식은 MAE 비율값이 4.8%, 석식은 MAE 비율값이 4.93%로 나타났다. 중식과 석식 모델이 우수한 성능을 보였다. 조식 모델의 예측 결과가 다른 모델에 예측 결과보다 오차가 큰 이유는 조식 인원 데이터의 추이가 다른 시간대 데이터보다 변동폭이 크기 때문에 복잡한 패턴을 학습하지 못하는 머신러닝 모델의 결과로 오차가 크게 나온 것 같다. 또 딥러닝 모델의 경우 충분한 데이터가 없으면 성능이 떨어진다. 이러한 문제들을 개선하기 위해 유의미한 양의 데이터 수집이 필요하다.

## V. Conclusions

본 논문에서는 이전 연구를 개선하여 2018년, 2019년, 2023년 학기 중 해당 날짜의 실제 식당 데이터와 기상청 날씨 데이터를 기반으로 조식, 중식, 석식 인원을 예측하는 모델을 개발하였다. 조식 모델은 XGBoost가 가장 좋은 성능을 보였고 중식 모델과 석식 모델은 Random Forest가 가장 좋은 성능을 보였다. 이 모델은 식당 운영자가 쉽게 사용할 수 있는 프로그램으로 구현되었다. 향후 연구 방향으로는 데이터 추가, 새로운 파생 변수 도입, 새로운 알고리즘 선택, 하이퍼파라미터 조정을 통해 모델의 성능을 더 향상시키고 프로그램의 UI를 사용자 친화적으로 개선하여 식당 운영자가 더 편리하게 이용할 수 있도록 연구를 진행할 예정이다.

## REFERENCES

- [1] Agnieszka Lasek, Nick Cercone, and Jim Saunders, "Restaurant Sales and Customer Demand Forecasting: Literature Survey and Categorization of Methods," *Smart City 360°*, pp.479-491, 2016.
- [2] J. Jeon, "Forecasting Models for Number of Persons for Meals in Large Food Service Place : Based on Actual Data of the S City Hall Restaurant for Employees," PhD thesis, Kyunghee University, 2019.
- [3] L. Chung, "The development of a forecasting model as a management strategy in university foodservices," Master's thesis, Yonsei University, 2001.
- [4] L. Chung, I.S Yang, and S.H. Baek, "Investigation on the Performance of the Forecasting Model in University Foodservice," *The Korean Journal of Nutrition*, Vol. 36, No. 9, pp.966-973, 2003.
- [5] H.J. Lim, "Comparative Assessment of Forecasting Models Applicable for Business and Industry Foodservice Operations," Master thesis, Yonsei University, 2008.
- [6] U. Kwon, "Analysis of priority factors in demand forecasting by cafeteria service providers : Focused on large-scale industrial cafeteria service providers," Master's thesis, Sogang University, 2020.
- [7] K.O. Cullen, L.W. Hoover, and A.N. Moore, "Menu Item Forecasting Systems in Hospital Foodservice," *Journal of the American Dietetic Association*, Vol. 73, No. 6, pp.640-646, Dec. 1978.
- [8] K. Jung, T. Jo, and K. Kim, "The Development of Prediction Models for the Number of People for Meal at University Cafeteria," *Proceedings of the Korean Society of Computer Information Conference*, Vol. 31, No. 2, pp.535-536, Jul., 2023.
- [9] KMA - Open MET Data Portal, <https://data.kma.go.kr/>
- [10] Y.J. Jang, H.S. Heo, B.J. Kim, S.K. Kim, G.M. Hong, and W.K. Lee, "The Change of The Average Discomfort Index from June to September during The Past 10 Years," *Journal of Climate Change Research*, Vol. 3, No. 2, pp.89-100. Jun. 2012.
- [11] E. C. Thom, "The Discomfort Index," *Weatherwise*. Vol. 12, No. 2, pp. 57-61, Apr. 1959. DOI:10.1080/00431672.1959.9926960
- [12] P. Schneider and F. Xhafa, "Anomaly Detection and Complex Event Processing over IoT Data Streams With Application to eHealth and Patient Data Monitoring," *Academic Press*, pp.59-60, 2022. DOI:10.1016/C2020-0-00589-X
- [13] Hodson, T. O.: Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not, *Geosci. Model Dev.*, 15, 5481-5487, <https://doi.org/10.5194/gmd-15-5481-2022>, 2022.
- [14] Chicco D, Warrens MJ, Jurman G. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Comput Sci.* 2021 Jul 5;7:e623. doi: 10.7717/peerj-cs.623. PMID: 34307865; PMCID: PMC8279135.
- [15] MLflow, <https://mlflow.org/>
- [16] FastAPI, <https://fastapi.tiangolo.com/ko/>
- [17] Kafka, <https://kafka.apache.org/>
- [18] Grafana, <https://grafana.com/>



## Authors



Gwangwon Jung received his B.S. degree in Computer Engineering from Mokpo National Maritime University, Korea, in 2024. Since April 2024, he has been working as an AI Research Engineer at StudioM Co., Ltd.

His research interests include Generative Multi-modal AI, Computer Vision AI, and AI-generated Content. He is particularly interested in utilizing AI to enhance the quality of media content for creative industries.



Keewon Kim received his M.S. and Ph.D. degrees in Computer Engineering from Kyungpook National University, Korea, in 2001 and 2006, respectively. He is currently an assistant professor in the department of

Computer Engineering, Mokpo National Maritime University. He is interested in information security, security protocol, VLSI, and big data analysis.