



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

박사학위논문

기계학습을 활용한 집단 급식소의
食數 豫測 모델 開發 연구

-S 시청 구내 직원 식당의 실데이터를 기반으로-

지도교수 권오병

경희대학교 대학원
경영학과

전 종 식

2019년 2월

기계학습을 활용한 집단 급식소의

食數 豫測 모델 開發 연구

-S 시청 구내 직원 식당의 실데이터를 기반으로-

지도교수 권오병

이 논문을 박사 학위논문으로 제출함

경희대학교 대학원

경영학과

전 종 식

2019년 2월

전종식의 경영학 박사학위 논문을 인준함

주심교수 서영호 ①인

부심교수 한명주 ①인

부심교수 김민용 ①인

부심교수 양희동 ①인

부심교수 권오병 ①인

경희대학교 대학원

2019년 2월

< 목 차 >

<국문 초록>	i
제1장 서 론	1
제1절 연구 배경 및 필요성	1
제2절 연구 목적	4
제2장 문헌 연구	6
제1절 예측 및 기계학습의 중요성	6
제2절 조리 외식 산업 수요 예측	9
제3절 집단 급식 관련 수요 예측	12
제4절 음식물 쓰레기 처리의 문제점	27
제5절 집단 급식소의 음식물 쓰레기	36
제6절 메뉴의 중요성과 메뉴 분류 체계	39
1) 메뉴의 개념과 메뉴 분석의 중요성	39
2) 통일 되지 않은 한식 메뉴 분류 체계	44
제7절 본 연구의 차별점	53
제3장 연구 방법	55
제1절 연구 모형	55
제2절 연구 대상	58
제3절 활용 데이터	59
제4절 구성 변수	62
제5절 구성 변수의 상세 내용	64
1) 식사 가능 인원	64
가) 식수 가능 인원 선정	64

나) 식수 가능 인원과 실제 식수 인원의 관계 분석	67
2) 시계열 인원 변수	69
가) 전주 식사 인원	72
(1) 전주 식사 인원 변수 생성	72
(2) 실제 식수 인원과 전주 식수 인원의 상관 관계 분석	74
나) 요일별 식수 인원 (요일 평균 변수)	75
(1) 요일별 식수 변수 생성	75
(2) 실제 식수 인원과 요일별 평균 식수 인원의 상관 관계 분석	77
다) 월별 식수 인원 (월 평균 변수)	78
(1) 월별 식수 인원 변수 생성	78
(2) 실제 식수 인원과 월별 평균 식수 인원의 상관 관계 분석	79
3) 메뉴 특성 변수	81
가) 한식 메뉴 분류 체계 수립	81
나) 한식 메뉴의 분류 기준	84
다) S시 집단 급식소 메뉴의 적용	90
라) 실제 식수 인원과 메뉴 특성 변수의 상관 관계 분석	92
4) 날씨 변수	96
5) 이벤트 여부 변수	101
6) 요일 및 계절 변수	102
7) 전날 혹은 다음날 공휴일 휴무일 여부 변수	103
8) 전후 연휴 및 연말 여부 변수	105
제4장 분석 예측 모델	107
제1절 분석 예측 모델 비교	107
1) 후진 단계적 선택법 (Backward Stepwise Selection)	111
2) LASSO	115
3) Ridge Regression	117
4) Random Forest	119
5) Bagging	122
6) Boosting	124
제2절 분석 예측 모델 결과 평가	126

제3절 분석 예측 최종 모델링	128
제4절 메뉴 개수 반영에 따른 식수 예측 모델링 비교	130
제5절 예측 최종 모델링 적용시 예상 효과	132
 제5장 결 론	 133
제1절 시사점	133
제2절 한계점	135
 <참 고 문 헌>	 136
 <감 사 의 글>	 148
 <ABSTRACT>	 149



〈표 차 례〉

〈표 2—1〉 단체 급식소에서의 식수 예측 관련 중요성	16
〈표 2—2〉 집단 급식소의 급식 생산 단계별 음식물 쓰레기 감량 방안	33
〈표 2—3〉 메뉴의 정의	41
〈표 2—4〉 「음식법 (饌法)」에 나오는 주식류의 종류, 재료, 빈도	46
〈표 3—1〉 식수 예측 모델링을 위해 사용한 공공 데이터	60
〈표 3—2〉 실제 식수 모델링 사용 변수 내용 요약	60
〈표 3—3〉 실제 식수 모델링을 위한 가용 데이터 요약	61
〈표 3—4〉 식수 가능 인원 일일 현황 파악표	64
〈표 3—5〉 요일별 식수 가능 인원 현황표	66
〈표 3—6〉 전주 식수 인원 생성표	72
〈표 3—7〉 요일별 식수(요일 평균) 인원	75
〈표 3—8〉 요일별 식수(요일 평균) 인원 생성표	76
〈표 3—9〉 월별 식수 (월 평균) 인원	78
〈표 3-10〉 월별 식수 (월 평균) 인원 생성표	78
〈표 3-11〉 단순 회귀 방식을 통한 인원 관련 변수의 설명력 요약	80
〈표 3-12〉 한식 메뉴의 듀이 십진법 분류법의 수정 개정안에 따른 분류 체계 요약	84
〈표 3-13〉 본 연구에서 적용한 한식 음식 메뉴의 분류 절차 요약	85
〈표 3-14〉 농식품 종합 정보 시스템의 식단 관리 메뉴젠의 간소화된 대분류	87
〈표 3-15〉 농식품 종합 정보 시스템의 식단 관리 메뉴젠의 간소화된 중분류	88
〈표 3-16〉 한식의 듀이 십진법 분류법 수정 개정안에 따른 메뉴의 분류	89
〈표 3-17〉 S시로부터 제공 받은 식단	90
〈표 3-18〉 단품 메뉴 특성 분류 예시	91
〈표 3-19〉 실제 식단 메뉴 분류 예시	92
〈표 3-20〉 주요 변수의 설명력 요약	95
〈표 3-21〉 기상청 데이터	96
〈표 3-22〉 날씨 데이터 전처리	98
〈표 3-23〉 비와 눈의 판단 기준	99
〈표 3-24〉 이상 기온의 판단 기준	99
〈표 3-25〉 날씨 데이터 전처리 결과 일부 예시	100

〈표 4—1〉 각 seed별 Backward Stepwise Selection에 의한 예측값과 실제값의 오차율	113
〈표 4—2〉 각 변수 별 분산 팽창 계수	114
〈표 4—3〉 LASSO 회귀 모델링 결과의 변수들의 계수값	115
〈표 4—4〉 각 seed별 LASSO에 의한 예측값과 실제값의 오차율	116
〈표 4—5〉 Ridge 회귀 모델링 결과의 변수들의 계수값	117
〈표 4—6〉 각 seed별 Ridge에 의한 예측값과 실제값의 오차율	118
〈표 4—7〉 각 seed별 Random Forest에 의한 예측값과 실제값의 오차율	121
〈표 4—8〉 각 seed별 Bagging에 의한 예측값과 실제값의 오차율	123
〈표 4—9〉 Boosting에 의한 영향력 있는 변수의 상대적 영향력	124
〈표 4-10〉 각 seed별 Boosting에 의한 예측값과 실제값의 오차율	125
〈표 4-11〉 각 seed별 모델링 기법에 의한 예측값과 실제값의 오차율 비교	126
〈표 4-12〉 상위 4개 모델링의 오차율 평균	128
〈표 4-13〉 최종 모델링 방법	128
〈표 4-14〉 3년간의 식단 메뉴 예시	130
〈표 4-15〉 메뉴 유형	130
〈표 4-16〉 메뉴 3개와 메뉴 전체에 대한 모델링의 오차율 결과	131
〈표 4-17〉 메뉴 개수에 따른 최종 모델링의 오차율 결과	131

〈그 립 차 례〉

〈그림 2—1〉 경기도 음식물류 폐기물의 처리 비율 추이	31
〈그림 3—1〉 식수 예측 모델 개발 연구 모형	55
〈그림 3—2〉 식수 예측 모델링 관련 적용 기법 및 최적의 예측 모델 생성	57
〈그림 3—3〉 식수 예측 모델링에 사용한 변수 요약	63
〈그림 3—4〉 요일별 평균 식수 가능 인원	66
〈그림 3—5〉 실제 식수 인원과 식수 가능 인원의 상관 관계	67
〈그림 3—6〉 실제 식수 인원과 전주 식수 인원의 상관 관계	73
〈그림 3—7〉 요일별 평균 식수 인원	76
〈그림 3—8〉 한식 메뉴의 듀이 십진법 분류법 수정 개정안을 기본으로 한 메뉴 분류 체계	83
〈그림 3—9〉 국립 농업 과학원의 농식품 종합 정보 시스템 식단 관리 메뉴젠	86
〈그림 3-10〉 이벤트 여부 변수 구분	101
〈그림 3-11〉 이벤트 여부 관련 변수값 부여 예시	102
〈그림 3-12〉 전날 혹은 다음날 휴무일 여부 구분	103
〈그림 3-13〉 전날 혹은 다음날 휴무일 관련 변수값 부여 예시	104
〈그림 3-14〉 연휴 전 혹은 연휴 후 여부	105
〈그림 3-15〉 연휴 전 혹은 연휴 후 관련 데이터 전처리 예시	106
〈그림 3-16〉 연말 관련 데이터 전처리	106
〈그림 4—1〉 Random Forest에 의한 변수별 중요도	120

〈국문 초록〉

기계 학습을 활용한 집단 급식소의

食數 豫測 모델 開發 연구

- S 시청 구내 직원 식당의 실데이터를 기반으로 -

전 종 식

경영학박사

경희대학교 일반대학원

지도교수 : 권 오 병

연구 배경

급식 업체에서 식수(食數)의 정확한 예측은 효율적인 급식 생산과 고객 만족을 위해 기본적으로 실행해야 하는 중요한 과정이며, 특히 잔반량 감소, 고객 불만 감소, 메뉴 품질 저하 방지, 비용 상승 예방에 필요하다.

집단 급식소에서 실시하던 기존의 식수 인원 예측은 주로 영양사가 본인의 경험을 토대로 한 직관에 의존하고 있는 상황이므로, 데이터 과학을 기반으로 한 식수 예측 모델링을 통해 예측력을 높여 집단 급식소의 잔반량 감소와 질 높은 고객 서비스 제공을 할 필요가 있었다.

연구의 필요성

잔반은 조리 과정에서 을 분류 기준으로 할 때 나오는 음식물 쓰레기로서, 음식을 만들기 위해 나오는 전처리 쓰레기, 미배식 음식, 먹고 남긴 음식 등 3가지로 나누어지며, 본 연

구는 잔반을 줄이는데 초점을 두고 있다.

많은 집단 급식소의 식수 예측은 전문 인력의 부족, 그리고 제한된 예산 등의 이유로 거의 경험에 의한 예측이 주를 이루고 있다. 그러나 이러한 상황에도 불구하고 국내외적으로 식수를 정확히 예측하여 수행하려는 연구가 많이 부족한 상황이고, 있다고 하더라도 예측 모델이 비교적 단순하여 실무적으로 적용할 수 있는 정도의 성과가 나오는 식수 예측 모델과는 거리가 먼 실정이다.

식수 예측은 제공되는 음식의 메뉴가 다양하고 한 식단에도 여러 반찬 등이 포함되어 있으며, 메뉴 이외에도 예측시 고려할 대상이 다양하여 매우 어려운 문제이다. 이에 본 연구는 실제 현장에서 사용되고 있는 메뉴와 메뉴 이외에 고려할 수 있는 다양한 요인을 모두 고려한 식수 인원 수요 예측 모델을 수립하여 공공 기관 집단 급식소의 효율적 운영을 꾀하고자 한다.

연구 문제

연구 대상이 되는 S시청의 집단 급식소인 구내 식당에서는 연간 110,900L의 잔반, 일평균 485L의 음식물 쓰레기가 발생하고 있어, S시청의 구내 식당에서는 점심 식사 인원 1,200명을 기본으로 요일에 따라 추가하거나 또는 적게 경험적으로 식수 예측을 하고 있어 데이터 기반의 예측 모델이 필요한 상황이다. 이에 S시청의 실데이터를 기반으로 집단 급식소의 잔반량 감소를 위해 기계 학습을 활용한 식수 예측 모델을 개발하고자 한다

문헌 연구

기계 학습을 활용한 식수 예측 모델링을 위해 기계 학습 및 기계 학습을 적용한 예측과 관련한 문헌을 살펴보고, 음식 서비스 분야 및 집단 급식과 관련한 수요 예측을 실시한 문헌을 통해 식수 예측의 중요성과 의의 및 한계점을 살펴 본다. 또한 음식물 쓰레기의 문제점은 처리 비용 뿐만 아니라 악취 및 민원 등의 사회적 비용이 발생하고 있어 음식

물 쓰레기를 줄이는 노력과 관련한 문헌도 살펴 본다.

분석 대상

식수 예측 모델링을 위해 본 연구는 S시 시청 구내 식당에서 제공한 메뉴, 식사 가능 인원 그리고 날씨 등에 대한 데이터 3년치 (2015년 5월 1일 ~ 2018년 4월 30일 기준)를 대상으로 연구를 진행하였다.

한식 분류는 조리 재료와 조리 방식에 따라 기본적인 분류를 진행할 수 있으므로 이러한 분류와 듀이 십진법 분류법에 의한 수정 개정안을 큰 틀로 하여 조리 방법과 재료에 의한 음식 분류를 본 연구에 적용하였다.

본 연구에서는 듀이 십진법 분류법에 의한 수정 개정안을 바탕으로 메뉴를 크게 주식, 부식 및 밑반찬, 후식 및 간식 등으로 나누고, 메뉴에 있는 개별 음식을 조리 재료와 조리 방법에 2가지로 분류하여, 1차적으로 조리 방법에 의한 분류를 진행하였으며, 2차적으로는 재료에 의한 분류를 하여 진행을 하였다.

예측 모델

식사 인원 예측에 영향을 미치는 요인 및 변수로 식사가능 인원, 시계열 인원 변수, 메뉴 특성 변수, 연휴 연말 변수, 이벤트 여부 변수, 요일 및 계절 변수, 전날과 다음날 휴일 여부 변수, 날씨 및 기온 변수 등의 8개 카테고리의 63개의 변수들을 사용하여 기계 학습을 활용한 1차 모델링, 최종 모델링 등 식수 예측 모델링을 2단계로 진행하였다.

첫째, 1차 모델링으로써, 741개의 관측치 (2015년 5월 일 ~ 2018년 4월 30일)를 사용하여 요일, 날씨, 기온, 메뉴 등 8가지 카테고리의 변수들을 대상으로 식수 인원 예측 모형을 개발했으며, 분석 기법은 다중 회귀 분석 중 Backwards Stepwise Regression,

능형 회귀, LASSO, Random Forest, Bagging, Boosting 등 회귀 분석 모델링과 기계 학습 기법을 적용하였다. 그리고 각 모델의 예측력을 비교 평가하여, 1차 모델링을 적용한 기법 중 오차율 상위 4개의 분석 기법을 선정하였다.

둘째, 최종 모델링을 진행하였다. 1차 모델링 단계에서 선정된 상위 4개의 모델링의 예측치 값들의 평균값을 최종 모델링의 예측 값으로 선정함으로써 데이터 예측의 안정성을 꾀했다.

결론적으로 하나의 기법에 의해 특정 모델링을 진행하는 경우 과적합 등의 사유로 향후 예측시 정확도가 낮아질 수 있는 경우가 발생할 수 있어, 1차 모델링의 결과를 기본으로 각각 개발된 모델에 대한 예측력을 비교 평가하고 평가된 모델을 기반으로 향후 미래의 데이터 예측의 안정성을 위해 예측력이 높은 최적의 모델을 생성 진행하였다.

결론

본 연구는 S시청 집단 급식소의 실데이터를 기반으로 식수 예측 정확도 향상에 유의한 영향을 미칠 수 있는 다양한 변수를 탐색하고 데이터 과학을 기반으로 머신 러닝 기법을 적용함으로써, 공공 기관 집단 급식소에 활용 가능한 새로운 식수 인원 예측 알고리즘을 국내 최초로 개발하여, 식수 예측 모델링을 제안하고 그 성능을 분석하였다.

본 연구는 2가지 측면에서 기존의 다른 논문들과 차별점이 있다고 본다.

첫째, 한식 메뉴 분류 체계를 학자와 기관 마다 다르게 사용하고 있는데, 이러한 한식 음식 메뉴 분류 체계의 통일되지 않고 복잡한 한식 분류 기법에서 탈피하여 본 연구에서 새로운 시각으로 적용한 한식 메뉴 분류 방법은 기존 연구에서 시도하지 않은 방법으로써 복잡한 한식 메뉴를 단순화하여 신뢰성과 확장성을 기반으로 다양한 식수 인원 예측의 모델링을 적용할 수 있었다.

둘째, S시의 집단 급식소를 대상으로 실데이터에 기계 학습 기법을 사용하여 여러 예측 모형들을 비교 및 결합함으로써 타 문헌에서는 보지 못한 실무에 적용할 수 있는 수준까지 오차율을 낮출 수 있었다.

결론적으로 본 연구의 식수 인원 예측을 통해 식수 인원에게 적합한 식재료의 사전 주문, 이에 따른 사전적인 식재료 처리, 그리고 더 나아가 적합한 식수 예측 인원에게 따른 조리 및 운영을 할 수 있도록 함으로써 미배식 잔반량을 줄이고 이에 따른 적합한 식수 예측의 비용 절감을 통해 식사의 질을 개선하는데 사용하여 이용자의 만족도를 높이고 음식물 쓰레기를 줄여 환경적 사회적 비용을 줄일 수 있는 선순환 구조의 기틀을 마련해줄 수 있는 모델링을 제시할 수 있었다.

또한 본 연구를 통해 앞으로 주로 현장 실무자들의 경험치를 기반으로 식수 예측을 하고 있는 학교, 관공서, 지자체 등의 집단 급식소에서 데이터 과학 기반의 적정 식수 예측 방법 및 기법의 확산에 기여할 수 있으며, 이번 식수 예측 모델링 연구를 기반으로 집단 급식소에서의 식수 인원 예측 관련하여 실무적인 향후 활용도를 높일 수 있는 기틀을 마련하였다고 할 수 있다.

핵심어 : R, 기계 학습, 잔반, 잔식, 집단 급식소, 식수 인원 예측 모델, 음식물 쓰레기

제 1 장 서 론

제 1 절 연구 배경 및 필요성

4차 산업 혁명 시대를 맞이하여 IT의 편리성, 속도, 정보의 양·질적 상승을 골자로 하는 정보 통신의 발달로 대규모의 정보 습득이 가능하게 되었다. 실질적으로 데이터 양이 기하 급수적으로 증가하고 있으며, 빅데이터를 기업 경영과 산업 발전에 활용하고자 하는 현장의 요구는 계속 강화되고 있다,

한편 기계 학습 응용 및 학습 알고리즘은 의치학, 체육, 경영, 유통, 회계학, 도로, 교통, 재난, 관광 등 다양한 분야에서 활용되고 있으며, 앞으로도 더 많은 분야와 영역에서 실 데이터를 활용하여 모델링을 하고 예측을 하는 기계 학습 기법의 적용이 이루어질 것이다. (이호현 외 2016)

그러나 아직 요식업 분야에서는 기계 학습 기술의 적용 및 활용이 아직 부족한 상태이며, 특히 집단 급식소에서의 식수 예측은 주로 경험에 의존하여 이루어지고 있는 실정이다.

집단 급식소의 효율적 운영을 위해 이를 운영하는 대기업에서는 그동안 자체적인 식수 예측 모델링으로 식수 인원을 추정한 것을 바탕으로 식단 기획 및 이에 대한 발주를 진행하여 운영해 오고 있다.

그러나 대기업에 의해 운영이 되고 있는 일부 집단 급식소를 제외한 많은 집단 급식소에서는 전문 인력의 부족, 제한된 예산 등의 이유로 체계적이고 과학적인 식수 예측이 이루어지지 않고 있으며, 부정확한 식수 인원 수요 예측으로 인해 여러 가지 부작용이 발생하고 있다.

현재 대다수의 많은 집단 급식소에서는 경험에 의한 잘못된 식수 인원 예측을 함으로써 잘못된 식수 인원 예측이 잘못된 발주 기획, 발주량의 수정을 통한 발주 수량 감소 및 추가 발주, 식재료비 관련 비용 증가 등 구매 계획의 차질, 식자재 저장 및 수급에 대한 문제 발생 야기 등의 문제로 이어져 비효율적인 인적 운영 및 효율성 저하, 생산 제공 음식에 대해 단편적이고 순간적인 대응으로 인해 민원이 야기 될 수 있는 조리 방법 및 잔식 발생, 음식 서비스 품질 저하 그리고 음식 품질이 발생되고, 이로 인한 배식 등으로 인해 민원 발생 및 생산 과부족으로 인한 비용 발생 그리고 고객 불만의 상승 등으로 이어지게 된다.

이러한 상황에도 불구하고 국내에서는 식수 예측 관련 연구가 식수 예측에 영향을 미치는 중요 변수가 무엇인지를 파악하는 정도의 이론적인 연구에 그치고 있으며, 식수 예측 관련하여 실무에 적용을 할 수 있을 정도의 예측 정확도를 보인 연구가 아직 부족한 실정이다.

또한 대형 공공 기관에서의 집단 급식소의 효율적 운영은 공적 가치 추구 측면에서도 매우 중요하다. 한 예로 시청과 같은 대형 공공 기관에서 집단 급식소에서 잘못된 수요 예측으로 대량의 음식 쓰레기가 방출되어 (예: S시청 연 약 11만 리터), 이로 인한 환경 오염과 세수 낭비가 발생 중이다.

식수 인원 예측의 어려운 점 중 하나는 예측시 고려할 대상의 다양성이다.

식수 인원 예측은 제공 되는 음식 의 메뉴가 다양하고 한 식단에도 여러 반찬이 포함되어 있으며, 식수에 영향을 미치는 변수들이 많은 복합 문제로서 예측의 어려움이 있다.

또한 우리 나라에서는 집단 급식소에서의 잔반에 대한 분석이나 총 잔식량에 영향을 미치는 잔식 발생 원인 분석, 잔식 감량화를 위한 피급식자의 기호도 조사, 음식물 쓰레기 관련 환경 오염 문제 등 음식물 쓰레기를 처리하는 방안 및 이와 관련한 전반적인 식생활 기반의 교육 등에 관한 이론적인 연구가 아직은 부족한 실정이다. (박세정 2017)

통상적으로 음식물 쓰레기는 크게 3가지로 분류가 되는데 음식물을 조리하기 전에 음식을 만들기 위해 전처리 과정에서 나타나는 음식물 쓰레기, 그리고 식수 예측을 잘못함으로써 나타나는 사용하지 않은 미 배식의 음식인 채로 버리게 되는 잔식, 그리고 식사 후에 사람들이 먹지 않아 일부가 남게 되어 버리게 되는 잔반 등이다. 이중 많은 집단 급식소에서 음식물 쓰레기 중 전처리 단계에서 나타나는 음식물 쓰레기와 식사 후에 나타나는 잔반은 운영자 입장에서는 관리를 할 수 있는 음식물 쓰레기이나, 식수 예측을 잘못하여 미배식 음식인 채로 버리게 되는 잔식은 영양사 등 운영자 측면에서는 항상 신경을 써서 해결해야 할 포인트라고 할 수 있다.

또한 이윤 추구하고 영양가 있는 선호 메뉴 사이에서 trade off가 존재하고 있고 식수 인원에 영향을 미치는 많은 환경 변수가 있어 위탁을 받아 운영하고 있는 집단 급식소에서는 이윤 뿐 만 아니라 요일, 휴일 여부, 연휴 및 휴가 시즌 여부, 식단 메뉴, 조리 인원 관련 노동력, 그리고 날씨 기후 등의 환경 변수 등 여러 가지 변수를 고려하여 진행하여야 하기 때문에 경험에 의한 식수 예측에는 많은 제약이 따른다.

이에 본 연구는 실제 데이터를 가지고 데이터 과학을 기반으로 한 식수 예측을 통해 잔반을 줄이고 공공 기관 집단 급식소의 효율적 운영을 위한 선제적 대응 방법을 구축하는데 기여하고자 한다.

제 2 절 연구 목적

본 연구의 목적은 집단 급식소에서 발생하는 잔반을 줄일 수 있는 데이터 과학을 기반으로 집단 급식소의 식수 인원 예측 기법을 제안하고 실제 데이터를 통해 그 기법의 활용 가능성을 보이는 것이다.

본 연구는 효율적인 식수 인원 예측을 통해 메뉴를 기획하는 단계와 음식 조리를 위한 재료 발주 단계에서 부터 적합한 식수를 예상할 수 있도록 하는데 기여하고자 하며, 공공 기관 등의 집단 급식소의 음식물 쓰레기인 잔반을 줄이고 데이터 과학에 기반한 식수 인원 예측을 통해 집단 급식소의 효율적 운영을 위한 선순환 구조의 선제적 대응 방법 구축에 기여하고자 한다.

또한 모집한 실데이터를 기반으로 회귀 분석 등을 통해 어떤 변수가 식수 예측에 미치는 영향을 파악하고, 기계 학습 기법을 활용하여 식수 인원 예측 모델링을 진행한다.

본 연구를 통해 현재 영양사, 조리사 등의 경험에 의해 진행이 되고 있는 대다수의 집단 급식소를 위해 식수 예측과 관련하여 기계 학습 기법을 도입하여 집단 급식소의 식수 인원 예측에 대한 변수 파악과 데이터 과학을 기반으로 한 식수 예측 모델링을 통해 식수 예측 알고리즘을 개발하여 진행한다.

또한 식수 인원 예측 모델링을 통해 조리 기획 단계에서 부터 조리 음식 과부족이 발생하지 않도록 발주 단계에서 부터 계획적인 구매가 이루어질 수 있도록 한다.

본 연구는 궁극적으로 데이터 과학 기반의 식수 인원 예측을 통해 기획 발주 및 구매를 바탕으로 효율적인 인력 관리 및 배치, 과학적인 음식 생산 관리 등의 효율적 조리 운영, 잔식이나 잔반에 따른 식자재 재료의 손실 및 비용 최소화, 음식 품질 발생 방지 등의 배식이 이루어지게 함으로써, 버려지는 음식물 비용 최소화, 고객 만족도 증대 등으로 이어지는 효과를 가져 올 수 있게 하도록 하여 단체 급식소의 효율적 운영을 위한 선순환 구조에 기여하고자 한다.

본 연구를 위해 기존 관련 문헌을 통해 식수 예측에 영향을 미치는 변수를 파악하였으며, 또한 식수 예측에 영향을 미치는 중요 변수들 관련한 실데이터를 S시 집단 급식소에 의뢰하였다.

그 결과 S시청 구내 식당에서 제공한 메뉴, 식사 가능 인원 등의 실데이터와 기상청의 날씨 등의 실제 데이터 3년치 (2015년 5월 1일 ~ 2018년 4월 30일 기준)를 대상으로 하여 중요 변수를 파악하였으며, 이를 바탕으로 식수 예측 모델링을 진행 하였다.

본 연구를 통해 집단 급식 식수 예측과 관련한 중요 변수 파악과 각 변수가 식수 인원 예측에 미치는 정도를 파악하며, 또한 효율적이고 신뢰할 수 있으며 확장 가능한 메뉴 분류 체계를 개발하며, 실제 식사 인원과 식수 예측 인원의 오차율을 최소화하는 기계 학습 등을 통한 데이터 과학 기반의 식수 예측 모델링과 관련한 학술 연구가 계속 심화될 수 있으며, 또한 연구 결과가 실무적으로 이어질 것을 예상되어지므로 학술적인 측면과 실무적인 측면에서 기여점이 있다고 본다.

특히 본 연구는 실무적 측면에서 집단 급식소에서의 정확한 식수 예측을 통해 미배식 음식 감소와 이를 통한 식사 제공 예산을 절약할 수 있을 뿐 만 아니라 절약된 예산을 가지고 식사 메뉴의 품질을 향상시킬 수 있으며, 식사 메뉴의 품질 향상을 통한 고객 만족도 증가 및 고객 불만 감소로 이어지는 선순환 구조를 만들 수 있는 시발점이라는 점에서 그 의의가 있을 것이다.

제 2 장 문헌 연구

제 1 절 예측 및 기계 학습의 중요성

예측은 과거로부터 발생한 데이터 및 이와 관련한 변수들의 알려진 과거 값에 기초로 하여 우리가 알고자 하는 미래 사건을 알고자 하는 것으로 정의 된다 (Makridakis, Wheelwright, and Hyndman, 1998).

또한 예측은 기업이나 단체 그리고 조직의 계획 뿐 만 아니라 운영 등과 같은 조직의 주요 관리 기능 중에서 모든 활동에 선행될 뿐 만 아니라 데이터를 기반으로 하기 때문에 실무적인 차원에서 중요한 의사 결정 등 경영 활동의 기초가 되는 중요한 활동이다. (Robbins, 2000)

예측은 수요 예측의 종류에 따라 크게 3가지로 구분을 하는데, 첫째, 기술 진보율의 예측을 하는 기술 예측, 둘째, 미래의 경제 상황에 대한 예측을 하는 경제 예측, 그리고 셋째, 미래의 어떤 특정의 일정한 기간 동안 기업의 서비스나 제품에 대한 예측인 수요 예측 등 3가지로 구분 된다.

통상적으로 조직에서 관리자가 어떠한 계획 활동을 제대로 수행하지 못하거나 혹은 효율적으로 수행하지 못하게 되는 경우 그 조직이 제공하는 제품 또는 서비스가 고객에게는 올바르게 인식되지 않고, 제대로 받아들여지지 않게 되므로 많은 갈등과 문제를 야기 하게 된다.

통상적으로 이러한 경우에 고객에 대한 적합한 수요 예측이 안되는 경우가 많아 여러 문제가 발생되므로, 적합한 수요 예측은 관리자가 제품이나 서비스를 효과적으로 계획하는데 필수적이라고 할 수 있으며, 그 중요성이 점점 부각되고 있다 (Kisang Ryu,

SooCheong (Shawn) Jang, Alfonso Sanchez, 2008).

그러므로 보다 정확한 수요 예측을 위해 과학적, 통계적 기법에 따라 체계적인 분석을 하는데, 일반적으로 수요 예측 방법을 선택하는 경우 수요 예측에 들어가는 비용, 수요 예측 결과의 정확성, 과거 데이터 혹은 과거 기록의 활용의 정도 및 적합성, 그리고 수요 예측을 하고자 하는 예측 기간, 소비 행동 유형 등 전반적인 사항을 고려하여 진행한다.

또한 이러한 수요 예측의 유형은 크게 정성적 기법의 수요 기법과 계량적 기법 등으로 나뉜다. 즉, 첫째, 개인의 성향이나 각자의 주관과 판단에 따라 혹은 여러 사람들의 의견에 따라 수요를 예측하는 정성적 기법의 수요 예측이 있고, 둘째, 과거의 역사적 수요에 입각하여 미래 수요를 예측 하는 시계열 분석 기법과 셋째, 수요와 긴밀히 관련되어 있는 변수들과 그리고 우리가 알고자 하는 수요와의 인과 관계를 분석하여 미래 수요를 예측하는 인과형 모형 예측 기법 등으로 이루어진 계량적 기법이 있다.

특히 수요 예측은 각종 생산 뿐 만 아니라 기업의 여러 활동을 위한 중요한 의사 결정에 기초 자료를 제공하며, 특히 계획 생산 및 구매에 중요한 역할을 하게 된다.

그리고 현업에서의 부정확한 수요 예측은 전반적인 운영에서 비용 측면에서 과다한 부담이 될 뿐만 아니라 이러한 부정확한 수요 예측으로 인해 발생된 여러 문제로 인해 조직원 간의 갈등이 야기됨으로써 이러한 갈등이 조직의 비효율적인 운영으로 이어지게 되어 여러 가지 나쁜 결정을 초래할 수 있다는 것을 인식하고 있다.

최근 예측력을 향상을 위해 기계 학습 (Machine Learning)이 관심의 대상이 되고 있다. 기계 학습은 “환경과의 상호 작용에 기반한 경험적인 데이터로부터 스스로 성능을 향상시키는 시스템을 연구하는 과학과 기술”이며, 스마트 디바이스의 확산과 쌍방향 통신의 활성화로 인해 데이터의 양이 방대해짐에 따라 기계 학습을 적용할 범위가 점차 확대되고 있다 (이호현 외 2016). 또한 향후 정확하면서도 저렴하고 빠른 정보 처리와 계산을 필요로 할 뿐만 아니라 데이터를 기반으로 한 예측 및 의사결정이 필요하기 때문에 기계 학습이 점점 더 각광받고 있다.

기계 학습에 적용 가능한 데이터의 양이 많아지고 컴퓨터의 속도와 기능 향상 및 기계 학습 기법의 적용 분야 확대에 따라 기존의 전통적인 통계적인 기법에서 탈피하여 실제로 업무에서 양산된 실데이터를 근간으로 분석을 하고 향후 일어날 일을 예측하는 등 많은 실제 분야에서 예측 알고리즘이 적용되고 있다.

딥러닝 등 기계 학습은 크게 첫째, 데이터 수집, 둘째, 알고리즘을 통한 데이터 학습, 셋째, 개발된 알고리즘에 대한 향후 발생한 미래의 데이터에 대한 유의미성 테스트 등의 3 단계로 이루어지고 있으며, 각 단계에서의 모델링과 알고리즘의 결합을 통해 성능을 향상시키고 보완시켜 목적에 맞게 진행을 하여 왔으며 이를 통해 실제 업무에 적용하고 있다 (이호현 외 2016).

그동안 기계 학습을 바탕으로 과거의 데이터로 파악한 패턴을 기반으로 미래를 예측하고, 발생 가능한 위험을 줄이고자 하는 많은 선제적 노력을 시도하여 왔으며, 또한 법, 의료, 제조 등 여러 분야에서 실질적인 기여를 하여 왔다.

오늘날과 같이 미래에 대한 불확실성이 가중되고 있으며, 미래에 발생하는 여러 가지 위험에 대한 기획 비용과 손실이 점점 커지고 있는 상황에서 예측은 미래에 일어나는 사건에 대한 선제적 대응에 필요한 작업이 되었다. 이는 사후적 대응 보다 미래의 일어날 일에 대해서 선제 대응을 함으로써 처리하는 비용이 작기 때문이다. 이로 인해 학계, 공공기관, 기업 뿐만 아니라 지자체 등 사회 곳곳에서 예측 기법은 많은 관심을 가지고 있다.

기계 학습에서의 예측력 강화를 위해서는 무엇 보다 적합한 데이터를 확보하는 것이 중요하다. 그리고 모델의 적합성을 위한 적합한 변수 확보 및 선정이 필요하며, 적합한 데이터 및 변수, 모델 기법, 해석력 등이 결합되어 이를 자동화 및 반복 프로세스를 통해 확장성을 가질 수 있도록 해야 한다. 또한 성능 개선을 위해 앙상블 모델링이 고려될 수도 있다 (SAS Institute, 2015 “Machine Learning: What is & Why is matters”).

본 연구에서도 데이터 과학의 식수 예측력 제고에 적합한 실데이터를 확보하고, 다양한 알고리즘을 대안으로서 고려하고 적합 모델링을 개발하는 방향으로 연구를 진행한다.

제 2 절 조리 외식 산업 수요 예측

수요 예측은 조리 외식 산업에서도 중요한 부분일 뿐 만 아니라, 특히 집단 급식소에서는 보다 정확한 식수 예측이 이루어져야 적절한 발주, 구매, 조리, 배식 등도 가능하므로, 정확한 식수 인원 예측은 집단 급식소의 선순환 실현에 가장 기초가 된다고 하여도 과언이 아니다.

조리 외식 산업에서의 수요 예측의 특징은 다음과 같다.

첫째, 조리 외식 산업은 먼저 일정한 시점이 지나면 버릴 수밖에 없는 제품의 소멸성이 강한 음식 자체의 본질적인 특성을 가진다 (Messersmith & Miller, 1991; Miller & Shanklin, 1988). 그러므로 많은 음식은 대부분 시간이 지남에 따라 부패하고 가격이 내려가게 되는 특징이 있어 유통 기한의 존재 및 음식의 상함으로 인한 발생하는 문제를 저감하기 위해 수요 예측이 절실하다.

둘째, 대부분의 음식 서비스 분야에서의 음식 아이템들은 서비스 직전에 만들어지고 있는 것이 많아 부정확한 수요 예측은 과잉 공급 혹은 과소 공급으로 이어지고 있으며, 사용되지 않고 남은 음식은 바로 비용으로 이어지게 되고 이로 인해 많은 문제가 야기되고 있다.

셋째, 통상적으로 바로 사용되지 않은 음식은 다음 식단으로 이어지는 경우가 많은데 이 경우 시간에 지남에 따라 음식의 질은 점점 더 떨어지게 되고 또한 버리지 않는 경우 사용되지 않은 음식은 보관으로 이어지게 되는 바 장기화된 보관에 의해 음식의 오염이 발생하는 원인이 되어 또 다른 문제를 야기하기도 한다. (Kisang Ryu, SooCheong (Shawn) Jang, Alfonso Sanchez, 2008).

마지막으로, 초과 수요 예측은 더 많은 노동력을 필요로 하게 되고 이로 인해 추가적인 인건비를 야기하게 되며, 수요에 부족한 예측은 고객 불만으로 이어져 민원을 야기하게 된다.

그러므로 조리 외식 분야에서는 실무적으로 적합하고 단순하고 시간을 절약할 수 있는 식수 인원 예측 서비스가 실질적으로 현장에서 매우 필요한 상황이다.

조리 외식 사업 분야에서 주로 사용되고 있는 수요 예측 방법은 첫째, 시계열 분석법으로 이는 시간이 경과함에 따라 특정 숫자나 변수의 변화의 추세, 트렌드, 그리고 경향 등을 분석하는 방법이 있다. (양일선 외 2011)

시계열 분석법은 과거 행동의 발생 유형이 미래에도 유효하다는 가정에 근거하여 사용하는 방법으로 과거의 매출이나 수량 자료로 부터 시간적인 추이나 경향을 우선적으로 파악하여 미래의 수요를 예측하는 기법으로 일 별, 주 별 그리고 월 별 등의 기준으로 혹은 분기 별, 반기 별 그리고 년도 별 등의 시간에 따라 축적된 자료를 활용한다.

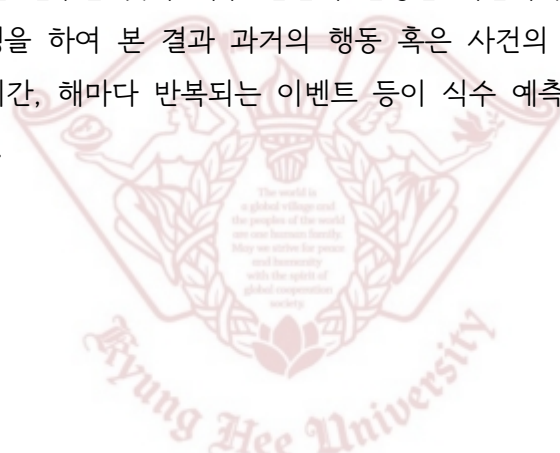
둘째, 다중 회귀 분석, 선형 회귀 분석 등과 같이 어떤 주어진 환경이나 조건 등에 대한 원인과 결과 등의 수학적 인과 결과를 나타내는 모델을 생성하여 예측을 하는 방법으로 인과형 예측법이나 전문가의 경험에 따른 주관적 판단에 의한 방법인 주관적 예측법이 그 예로서, 인과형 모형 기법의 수요 예측은 우리가 예측하고자 하는 수요를 종속 변수로, 그리고 수요와 관련되어 영향을 미치는 여러 요인들을 독립 변수로 놓고 이러한 변수들 간의 관계를 정리하여 모델링을 통해 여러 가지 모형으로 파악하여 수요를 예측한다.

본 연구에서도 또한 어떠한 변수들이 실제 식수 인원에게 얼마나 영향을 미치는지를 파악하고자 하였으며, 이를 통해 과거 식수 예측에 경험적으로 알고 있었던 많은 현상을 설명하는 변수에 대해 보다 구체적으로 확인을 할 수 있었으며, 보다 적합한 식수 예측을 위해 특정 변수에 대한 추가적인 필요성을 발견할 수 있었다.

셋째, 복잡한 사회 문제를 풀기 위한 통계 등을 응용한 기계 학습 기법이나 딥러닝 기법 등이 있다.

한편 조리 외식 산업에 있어서 음식 소비량 예측에 있어 계절에 따른 패턴이 매해 일정하다고 할 때, 계절성을 반영한 데이터가 더욱 정확한 예측에 도움 되기도 하나 (Farnum & Stanton 1989), 이는 주별, 혹은 월별 혹은 일별로 아주 특별한 패턴을 보이는 경우에만 적용될 수 있다는 한계점이 있어, 연구의 목적이나 성격에 따라 적용 가능성의 차이가 있다.

본 연구에서도 식수 예측 모델링을 진행하는 데 있어서 3년치의 데이터를 확보하여 주별, 월 별 등의 시계열 변수를 찾아 식수 인원에 일정한 패턴이 있음을 확인하고자 하였으며, 실제적으로 진행을 하여 본 결과 과거의 행동 혹은 사건의 발생 유형이 월 별, 주별 혹은 연말, 연휴 기간, 해마다 반복되는 이벤트 등이 식수 예측에 중요 변수로 작용함을 확인 할 수 있었다.



제 3 절 집단 급식 관련 수요 예측

집단 급식 혹은 단체 급식이라 함은 50인 이상의 특정의 다수인에게 학교나 병원, 군대, 단체, 공공 기관, 지자체, 산업체 등에서 일시적으로 식사를 제공하는 것이 아니라 지속적으로 식사를 제공하는 것을 말한다 (조희숙, 1998)

집단 대형 급식에 관한 연구는 제공되는 식사 품질에 대한 연구, 메뉴의 만족도에 대한 연구, 급식 서비스 요인에 대한 연구, 식수 예측에 대한 연구 등으로 이루어지고 있다.

첫째, 식사 품질에 대한 연구로서 조직의 근로자에게 제공하는 웰빙 문화의 특성을 유형화하고 개념을 정의하고 수익 분석을 하기도 하였으며 (권수연 외 2010), 근로자에게 제공되는 점심 식단을 조사하고 영양 제공량을 분석하기도 하였다 (이슬 2010).

둘째, 메뉴 만족도에 대한 연구로서 근로자의 급식 메뉴에 대한 만족도를 측정하는 연구가 있었고 (백옥희 외 2007), 음식의 신선도, 음식의 영양적 균형, 선호 메뉴 등에 대해 나타나는 만족 요소 뿐 만 아니라 여러 불만족 요소를 지적하기도 하였다.

셋째, 급식 서비스 요인에 대해서는 음식 품질, 공조 환경, 인적 서비스, 청결성, 공간 배치 및 기능, 메뉴 다양성 등이 이용자 만족도에 영향을 주고 있으며, 집단 급식에 영향을 미치는 요인으로 음식의 메뉴, 음식의 가격, 위생 요인, 식당 시설 등이 중요한 요인으로 지적되기도 하였으며, 고객의 만족도 영향을 주는 요인으로는 음식의 가격과 맛이 중요한 것으로 파악되기도 하였다 (이은용 외 2007 ; 오지연 2010).

또한 대형 급식에 있어서 고객들이 메뉴를 선택함에 있어 음식의 맛, 음식과 식기 위생, 식사 가격을 중요한 요소로 고려하고 있으며, 외식 산업 발달로 근로자들은 다양한 외식 문화를 접하는 기회가 늘어나고 소득 증대에 따라 더 높은 수준의 급식 서비스를 제공받기를 원한다는 것을 밝힌 연구도 있다. (이해영 외 2004)

마지막으로 식수 예측에 대한 연구는 그 중요성에도 불구하고 많이 이루어지지 않고 있다. 예를 들어 대형 급식소의 배식 방법에 따른 음식 잔식량에 대한 연구에서 집단 급식소에서 정량 배식의 잔식량이 자율 배식의 잔식량 보다 높다는 연구가 있기도 하였으나 (김혜영 2000), 집단 급식소에서의 식수 인원에 대한 예측 연구는 활발한 연구가 이루어 지기를 바라는 기대와는 달리 소수 연구만이 이루어졌다 (정라나, 2001; 임혜진 2008).

통상적으로 집단 급식소는 많은 사람을 상대로 하여 음식 서비스를 제공해야 하기 때문에 항상 신경을 써야 하고, 식사 시간에 사람이 한꺼번에 몰리는 특성 때문에 대량의 음식을 아주 짧은 시간 내에 조리해내야 해야 하는 특징이 있다.

특히 면역력이 약한 환자가 있는 병원 같은 곳에서의 집단 급식은 환자의 특성에 맞는 다양한 식사가 제공 되어야 하는 집단 급식의 특이성 속에서 대략 1시간 이내에 짧은 시간에 정확하게 급식이 이루어 져야 하기 때문에, 이로 야기되는 음식 서비스 품질의 부족은 심각한 고객 불만으로 이어지고 병원 운영상에 부정적 영향을 줄 수 있다 (Magatet & Beatrice, 1976; Cullen et al., 1978).

그러나 집단 급식소의 경우 식수 대상이 거의 고정적이고 예측 불가능한 외부 인원을 통제할 수 있어 식수 인원을 경험적으로 예측할 수 있으므로 전체 잔반을 합리적으로 줄이기에 적합하다 (전무영, 민혜선 1998).

이러한 특성 때문에 집단 급식소에서의 식수 예측은 양질의 데이터를 잘 관리하여 진행이 되는 경우 정확도 높은 식수 예측 모델링이 가능하다. 더욱이 최근 기계 학습 등 다양한 예측 기법의 발달로 인해 영양사, 조리사 등 현장에서도 수요 예측의 필요성에 대해 관심을 가지기 시작하였고, 운영자 입장에서 식수 예측 모델링에 대해 많은 필요성을 가지고 있는 상황이다.

이러한 상황 속에서 집단 급식의 시장 규모는 나날이 증가하고 있는데, 2007년 기준으로 약 7조 4000억 원으로 시장 규모를 추정하고 있고, 이중에서 위탁 운영하는 비율은 약

50% 수준인 3조 7000억 원으로 추정하고 있다 (한국 외식 정보, 2009).

또한 2012년 기준으로 집단 급식의 시장이 약 9조 7000억 원으로 확대 되었으며, 이 중 산업체 급식이 차지하는 비중이 제일 크다고 할 수 있다 (한국 외식 연감, 2013).

운영적인 측면에서 볼 때 외부에 운영을 맡겨서 진행하는 위탁 급식이 차지하는 금액은 4조 6000억 원으로 추정하고 있으며, 이중 약 3조 1000 억원의 시장을 대기업을 포함한 상위 9개 사가 차지하고 있다고 분석하고 있다 (이현주, 2013).

위탁 급식이 전체 집단 급식 시장의 거의 50%를 차지하고 있는 상황에 대기업을 포함한 상위 9개사가 차지하는 비중이 70%에 달하는데 이 상위 9개사를 제외한 곳이 운영하고 있는 집단 급식소는 거의 경험에 의해서 식수 예측을 진행하고 있다고 해도 과언이 아니라고 할 수 있다.

집단 급식 시장은 규모가 커질수록 개인의 기호 성향에 대한 배려는 낮아지고 급식 가격에 맞추어 질이 저하되는 경향도 나타나고 있기도 하며, 이로 인해 여러 부작용이 발생하고 있어 실질적으로 직원이나 구성원들에게 제공되는 식사 품질의 향상과 메뉴의 다양화는 매우 중요하다고 할 수 있다.

또한 집단 급식 시장을 운영하는 운영자 입장에는 매일 식사를 제공해야 하기 때문에 실제 집단 급식소 현장에서 일하는 근로자들의 근로 강도와 노동력을 고려하여 메뉴를 운영하고 있으며, 집단 급식소의 위탁 급식 운영의 경우 우선적으로 가격에 민감하게 반응하여 식단을 구성하기 때문에 식사 품질의 향상과 메뉴에 한계점이 노출되는 경향이 있다고 할 수 있다.

맛별이 부부의 증가, 다양한 사회 활동의 증가 등으로 인해 가정 외에 외부에서 식생활이 증가함에 따라 집단 급식소는 계속적으로 점점 늘어나는 추세이나, 실질적으로 집단 급식에 대해서는 양적인 면에서 팽창함에 따라 식단 영양가의 산출 오류, 식수 대상에 대한 기호 조사 미흡, 그리고 대규모 인원이 한꺼번에 물리는 것에 대해 대처할 수 있도록 하

는 대량 조리 시스템 미비 등으로 인해 영양상의 문제 뿐 만 아니라 서비스 전반에 대한 여러 가지 문제가 제기되어 왔다.

직원이나 근로자 등 식수 대상자들이 가지는 가정식에 대한 향수와 단일식단에 대한 불만 등이 여러 가지 민원 등으로 이어지고, 또한 제공되는 식단의 열량 부족 내지는 식단에 대한 영양에 대한 무지 그리고 자율 배식으로 인한 제공 식단이 과식의 원인이 되기도 한다.

집에서 먹는 음식과 달리 집단 급식소에서는 음식에 대해서는 직원이나 근로자 등 식수 대상자들이 집에서 먹은 음식과 비교 등으로 인해 다량 조리로 인한 시각적 후각적 식욕 저하 등 심리적으로나 정서적인 면에서 부정적인 영향을 끼치는 문제점이 있어 왔다.

따라서, 집단 급식소에는 많은 사람들이 몰리는 점심이나 저녁 식사 시간 대에 적시로 제공하는 음식 서비스의 대량 조리를 위해 이에 맞는 효과적인 조리 설비를 가지고 있어야 하며, 적시에 설비 기기를 이용하여 제 때 음식 서비스를 제공할 수 있어야 하고 또한 이를 식수 대상자들에게 적절한 배식 방법에 따라 효과적으로 나누어 줄 수 있는 체계적인 시설을 갖추고 있어야 한다 (Spears, 2000).

또한 이런 시설을 갖추고 있다고 하더라도 집단 급식소 내에서의 조리 종업원의 실수로 혹은 고의로 식기류의 세척 및 살균이 불충분 등으로 인한 유해한 세균의 침투 등으로 인한 위생상의 문제가 항상 내재 되어 있다.

그리고 지속적인 물가 상승으로 인해 질 높은 재료비 확보 부족, 그리고 조리 및 운영 관련 인건비를 비용 상의 문제점으로 과도하게 절약시 육체 노동이 많은 집단 급식소 내에서의 조리의 질이 저하되고 음식물 쓰레기가 증가하는 등 이로 인해 민원이 야기되거나 서비스 불만족으로 이어져 운영상의 많은 문제점이 노출될 수 있다.

집단 급식소에서 식수 예측이 과학적이고 체계적인 방법에 의해 운영되지 않는 경우 많은 문제점들을 야기하고 있다.

앞서 설명을 한 대로 현재 많은 단체 급식소에서는 경험에 의한 잘못된 식수 수요 예측 결과로 인해 발주량 수정, 추가 발주 발생, 식재료비 관리 통제 미흡 등의 발주로 인해 구매 계획의 차질, 식자재 저장시 그리고 식재료 수급 관련한 문제가 발생하는 등의 문제로 이어지고 있다.

이러한 문제점 및 현상은 집단 급식소에서의 효율적인 인력 운영 저하, 생산량 과부족에 대해 단편적인 대응 등에 따른 조리와 잔식 발생 그리고 저하된 품질 및 음식물 품질 발생 등으로 이어지게 되고 더 나아가 과잉 생산에 의한 잘못된 배식 혹은 제공 음식 부족에 따른 비용의 발생 그리고 고객 불만의 상승 등으로 이어지는 상황이다. (양일선 외 2011)

식수 수요 예측의 의의와 결과를 나타내면 표2-1과 같다. 표 2-1에서 보듯이 잘못된 식수 예측은 옷의 첫 단추가 잘 못 끼워지면 처음 부터 다시 끼워야 하는 것처럼 지속적인 문제를 야기 시키므로 식수 가능 인원과 잔식의 양과 잔식에 영향을 미치는 요인들을 파악하여 식수 인원의 수요 예측을 하는 것이 무엇 보다 중요하다고 할 수 있다.

표 2-1 단체 급식소에서의 식수 예측 관련 중요성

빛나간 식수 수요 예측 결과		현장에 맞는 성공적인 식수 인원 수요 예측 결과
<ul style="list-style-type: none"> 발주 음식 수정 및 추가 발주 식재료비 관리 통제 미흡 	발주	<ul style="list-style-type: none"> 발주시 발주량 및 메뉴 변경의 비효율성 감소
↓	↓	↓
<ul style="list-style-type: none"> 구매 계획의 차질 식재료 수급 문제 발생 및 관련 저장 문제 발생 	구매	<ul style="list-style-type: none"> 계획적인 구매 가능 식재료비 낭비 최소화
↓	↓	↓
<ul style="list-style-type: none"> 효율적인 인력 운영 저하 음식 생산 관련 과부족에 대해 단편적인 대응 	조리	<ul style="list-style-type: none"> 효율적인 인력 관리 및 배치 과학적인 음식 생산 관리
↓	↓	↓

<ul style="list-style-type: none"> • 잔식 발생 • 저하된 품질 및 음식 품질 발생 	배식	<ul style="list-style-type: none"> • 잔식 및 잔반 발생에 의한 식자재 관련 비용 및 손실 최소화 • 음식 품질 발생 방지
↓	↓	↓
<ul style="list-style-type: none"> • 과잉 생산 음식 또는 부족 생산에 따른 손실과 비용 발생 • 고객 불만의 상승 	효과	<ul style="list-style-type: none"> • 집단 급식 현장에 알맞은 음식 공급 따른 비용 최소화 • 고객 만족도 상승 및 증대

일반적으로 음식은 미리 만들어 지는 것이 아니라 먹기 직전에 만들어지고 먹고 난 후 나머지 음식은 부패하는 특성이 있고 다른 산업에 비해서 상대적으로 유통 기한이 짧기 때문에 집단 급식소에서 음식 서비스에 대한 수요 예측은 더 더욱 중요하다고 할 수 있다.

그러므로 음식의 특성상 기한이 일정 시점이 지난 음식은 음식의 질이 떨어질 수 밖에 없는 상황이고, 시간이 지남에 따라 부패하거나 상하기 쉬운 식자재를 많이 사용하는 단체 급식소 등에서는 얼마나 식수 인원의 수요 예측을 얼마나 잘하느냐가 매우 중요한 점으로 인식되고 있다. (정라나, 양일선 외 2003 ; Davis & Berger 1989)

또한 이러한 짧은 유통 기한의 음식의 특징 때문에 집단 급식소에서 유통 기한에 임박한 남은 식재료의 사용은 집단 식중독 등 대형 민원을 야기 시키면서 조직원 및 구성원들에게 심각한 후유증을 동반하는 위험을 야기 시킬 수 있다.

수요 예측은 미래의 요구를 예측하기 위해 체계적 방법에 의해 과거의 정보를 이용하는 기술로서 잘못된 수요 예측은 초과 생산 혹은 생산 부족을 초래하여 손실을 유발하게 되며, 생산량, 과잉생산 및 부족생산에 대한 비용, 종업원의 사기, 관리자의 자신감, 고객 만족 등에 영향을 미친다고 할 수 있다 (Messer smith & Miller, 1992 ; Harris, 1995).

집단 급식소에서 수요 예측이 잘못되면 앞에서 설명한 바와 같이 여러 가지 문제점들이 나타나면서 급식의 악순환 고리로 연결되어질 가능성이 높아진다.

수요가 많은 것으로 잘못 예측하는 경우 이는 식재료 낭비로 인한 비용 손실로 이어지고, 수요가 적은 것으로 잘못 예측하는 경우 이는 식사 부족으로 인해 고객 불만족, 종업원의 사기 저하 등 부정적인 영향을 미치게 되고 급식 생산 비용은 부족 생산일 때 더 높아지게 된다. (정라나 외 2003; Dougherty , 1984; Miller & Shanklin , 1988a ; Miller et al., 1991c; Messersmith & Miller, 1992; Spear, 2000).

그러므로 집단 급식소에서 식수 관련하여 잘못된 수요 예측을 통해 더 많은 양의 식사를 준비하게 되는 것은 미배식 잔반 발생으로 인한 비용 낭비, 제한된 비용에서 과다 비용 지출로 인한 품질 하락, 현금 유동성 저하로 이어지게 된다

그리고 이와 반대로 예상 인원을 적게 산정함으로써 실질 인원 보다 더 적은 양의 식사를 준비하게 되는 것은 고객 불만 초래, 긴급한 추가 발주, 음식 부족으로 인한 다른 음식 서비스로 이어져 고객의 불만족 및 원가 상승에 대한 부담이 가중되게 된다고 할 수 있다.

현실적으로 음식 서비스 산업은 다른 산업과 달리 기본적으로 노동 집약적인 특성이 강하다. 이는 결론적으로 집단 급식소에서의 식수 인원 에 대한 적합한 수요 예측이 제공되는 급식 서비스의 질과 강한 상관을 가지고 있다는 것을 의미하며, 음식 서비스 업종에서의 수요 예측이 잘 이루어져야 이에 맞는 효율적인 인력 배치가 가능하고 또한 이는 결국 인건비 절감 등의 효과로 이루어지게 된다고 할 수 있다. (Davis & Berger 1989)

그러므로 식수 예측은 적절한 식수 예측 알고리즘의 개발을 통해 집단 급식소의 미배식 잔반의 감소를 통한 원가 절감 그리고 조리 작업 업무의 효율화 등의 효과가 있다고 할 수 있을 뿐 만 아니라, 시설 투자 및 운영에 있어서도 많은 긍정적인 영향을 미친다고 할 수 있다. (정라나 2001)

또한 집단 급식소에서 초기에는 시설 투자를 위해 많은 투자와 비용이 들 뿐 만 아니라 수요 예측이 제대로 되지 않은 사항에서는 시설 투자를 진행하는 경우 시설의 낭비로

이어 진다고 할 수 있다. (정라나 외 2003)

이런 음식 서비스 산업에서의 식수 인원에 대한 수요 예측의 중요성 때문에 1960년 대부터 수요 예측을 위한 다양한 방법이 그동안 지속적으로 개발이 되어 왔다.

또한 정확한 식수 수요 예측 일수록 노동 비용과 식재료비가 절약되므로 식수 예측에 대한 정확성을 높이기 위해 다양한 방법들이 고안되고 개선을 위한 노력이 같이 병행됨으로써 오차를 줄이기 위해 많은 노력을 하여 왔다.

1970년대에 들어서서 Missouri-Columbia 대학의 있는 영양 및 균형식 (Nutrition and Dietetics) 건강 과학 연구소 (Health Sciences Centers)에서 음식 관련 서비스 수요에 대한 예측 연구가 진행되었다.

이 연구 센터에서는 컴퓨터를 활용하여 메뉴에 있는 음식 아이টে에 대한 수요를 예측하기 위해서 식사량의 패턴을 분석하고 또한 메뉴에 따라 달라지는 음식 수요의 측정 방법을 연구하였으며, 그리고 음식 서비스 관련한 수요 예측 시스템 개발, 그리고 이에 대한 적용 및 관리 운영에 관한 연구가 집중적으로 이루어 지기도 하였다 (Chandler et al., 1982 ; Dougherty, 1984 ; 정라나 외 2003).

또한 식단 메뉴의 아이টে에 대한 통계적 예측을 위해 박스-제킨스 (Box-Jenkins) 예측 모델이 일일이 손으로 하는 작업을 해서 식단에 대한 식수 인원을 예측 하는 데 적용될 수 있다는 것을 증명하기도 하였다. (Wood 1977)

그 후 병원의 급식 서비스 개선을 위해 연구가 진행되었는데, 환자 식수 인원 예측, 식사 종류별 비율 추정, 메뉴에 따른 수요 결정의 3단계로 구성하여 병원 급식 서비스 개선을 위한 다양한 식사 메뉴 개발의 병원 식수 예측 시스템을 개발하여 손으로 방대한 정보를 처리해야 하는 수요 예측 업무의 제한점을 지적하기도 하였다. (Messersmith 1978).

그리고 컴퓨터를 활용하여 많은 자료를 정확히 분석할 수 있으며 또한 자료 분석을 위한

시간 부족을 메꿀 수 있다고 하여 컴퓨터를 이용한 식사 메뉴에 대한 수요 예측 기술을 개발했다 (Messersmith 1978).

또한 병원 환자에게 제공되는 하루 세 번 제공되는 식사에 대한 환자 조사를 그래픽 및 분산 분석을 사용하여 제시하였으며, 제공되는 식사에 대한 신뢰할 수 있는 패턴이 환자 조사 데이터에서 확인하여 이를 수학적 예측 모델의 수용이 가능하게 하도록 하기도 하였으며, 식사 종류별 비율 측정의 단계를 제외하고, 환자 식수에 기초 했을 때, 예측 모델도 훨씬 간단하면서 정확도 또한 기대 범위에서 크게 벗어나지 않는다고 하기도 하였다 (Chandler et al 1982)

또한 병원의 환자 급식에 대한 연구가 지속되었는데, 2개 메뉴의 경우와 3개의 메뉴 항목에 대한 예측 시스템의 효율성을 예측 값을 통해 비교했다. 이를 비교하는 예측 관련 각각의 시스템에는 적응 지수 평활과 Box-Jenkins 모델이라는 두가지 예측 모델이 사용되기도 하였으며, 이 시스템은 예상 오류 비용 측면에서 비교하여 그 때까지 가장 복잡하지 않은 연구를 하였다 (Cullen et al 1978). 이후에도 병원 급식에서의 식수 인원에 대한 수요 예측에 관한 연구는 계속 되었다.

앞선 설명에서 보듯이 음식 서비스 분야에서 식수 예측 모델이 연구되고 제시되었지만 병원 환자 급식 서비스 등 매우 제한적인 범위에서 진행이 되는 한계점을 가지고 있어 널리 보급이 되지 못하였다. (Finley & Kim, 1986)

그러나 실질적으로 집단 급식소를 운영하고 있는 급식 운영자들은 식수 인원에 관한 수요 예측에 대한 필요성은 깊이 인식하고 있었으며 (Miller & Shanklin , 1988a ; Repko & Miller , 1990; Miller , 1991b), 이런 필요성에 따라 음식 서비스의 수요 예측 관련 모델이 개발되는 데 있어서 예측 모델이 복잡하다고 할지라도 이것이 수요를 항상 정확하게 예측을 하는 것이 아니라 오히려 잘못될 수 있다는 것을 지적한 연구자들도 많았다. (Cullen et al., 1978; Miller et al., 1991c; Armstrong, 1986; Georgeoff & Murdik, 1986)

병원에서 제공하는 급식 서비스에서 벗어나 병원 이외 다른 음식 서비스 분야로 확장하는 식수 인원 예측에 대한 연구가 1990년 대 부터 시작이 되었다.

그 예로 학교 급식에 적용할 수 있는 2가지 간단한 예측 모델 (이동 평균법 · 지수 평활법)을 제시하기도 하였는데 (Miller 등, 1991c), 이 연구에서는 발생하는 계절 영향 오차에 대해 단순 회귀 분석을 이용하여 진행을 한 초보적인 단계였다.

그 외에도 계속하여 급식 산업에서 사용되고 있는 식수 수요 예측 기법에 대한 비교 연구들이 수행되기도 하였다. (Miller, 1990 ; Miller et al., 1993 ; Liu & Ridgway, 1995; Sanchez et al., 1995; Pickert & Miller, 1996)

과거의 데이터를 바탕으로 식수 예측을 하는 것은 병원, 학교, 산업체, 관공서, 공공 기관 등 각 집단 급식소의 유형과 특성에 따라 식수 예측에 영향을 주는 요인이 다르기 때문에 식수 예측 기법을 다르게 적용될 수 있다고 하였다. (Cullen et al, 1978 ; Lin & Vassar ; Messersmith et al, 1978 ; 1992; Miller et al. 1991b)

식수 예측을 하기 위해서는 식수에 영향을 주는 요소가 급식을 제공하는 집단 급식소의 환경에 따라 조금씩 달라지는 데 집단 급식소의 식수에 영향을 주는 요소로 계절적인 영향이 존재하는 것은 아니라는 것을 밝히기도 하였고 (Miller 등, 1991c), 이외에도 수요에 영향을 주는 여러 요인에 대한 연구가 있었다.

메뉴 선호도에 대한 심리학적 요인에 대한 연구(Scrimshaw, 1969; Tanner, 1976; Levitt, 1967; Gatsos & Watkins, 1998)와 학교와 같은 집단 급식소 처럼 일정한 패턴을 보이는 학기제에 따른 영향을 분석한 연구들이 계속 되어 진행이 되기도 하였다. (Haach, 1970; Zifferblatt, et al., 1980)

식수 수요를 예측하기 위한 기법은 점차적으로 그 종류가 다양하게 되고 영향 요인 또한 여러 가지가 있지만 실질적으로 수요 예측을 정확하게 하기 위해서는 무엇보다 더 중요한 것은 메뉴와 식수 인원 예측에 미치는 변수들에 대한 과거 데이터 그리고 식수와 관련된

과거의 자료를 잘 기록하고 보관하는 것이 매우 중요하다고 지적을 하기도 하였다.

(Mackle & David, 1976; Dougherty, 1984; Spears , 2000)

시대에 따라 아무리 통계 기법이 발달하고 기계학습 기법과 딥러닝 기법을 사용한다고 하더라도 오차율을 줄이기 위해서는 연구자가 정확한 예측을 하기 위해서는 진행할 수 있는 데이터가 무엇 보다 더 중요하다고 할 수 있다.

이는 실제로 식수 예측 진행을 해보면 쉽게 알 수 있듯이 실질적으로 식수 인원 수요 예측을 진행하는 데 있어서 과거 자료의 부족은 통계 분석 혹은 기계 학습을 진행하는 경우 식수 예측 관련 오차가 크게 발생 하여 식수 예측에 대한 신뢰도를 떨어뜨릴 수 있기 때문이다 (이순목, 1995; Adebajo, 2000).

또한 식수 예측 관련하여 다른 산업과 달리 급식 산업의 특성 중 쉽게 상하는 식재료를 많이 쓰기 때문에 상대적으로 음식 재료의 재고 유효 기간이 짧아 음식물 잔반이 많이 발생할 수 있는 특성을 보완하는 노력을 하였으며, 이러한 맥락 상에서 시간 대 별 수요 예측을 연구한 사례도 있었으며 (Adebajo, 2000), 또한 식사 시간 대 몰려 드는 사람의 좀 더 정확한 예측을 위해 짧은 시간 단위 (예 : 15분 단위)의 작업 측정을 통해 정확한 수요 예측과 인력 배치를 위한 연구를 하기도 하였다 (Ansel et al 1999),

이에 대한 대표적인 연구 사례로 미국의 fast food 업체인 Taco bell의 경우, 하루 매출 중 점심 시간인 11시부터 14시까지 일일 매출의 52%를 차지하고 있어, 노무비가 매출의 30%를 차지하고 있어 노동력 배치가 중요 현안이었기 때문에 이를 위해 15분 단위로 고객 수를 측정하였고, 이를 적합한 식수 예측과 노동력 배치에 활용하여 1996년에 1,640만 달러의 비용을 절감할 수 있었다 (Hueter & Swart , 1998).

이는 놀라운 결과였으며, 이렇게 까지 할 수 있었던 것은 문제에 대한 해결점을 찾고 싶었던 Taco Bell과 이와 관련한 과거 데이터의 준비가 무엇보다 더 중요하다는 것을 깨닫게 해주는 좋은 사례였다고 할 수 있다.

이 외에도 많은 연구자들에 의해 음식 서비스 관련 인력 배치와 식수 수요 예측 간의 상관 관계 연구가 시도되었다. (Edwards & Ingram, 1995; DeFranco, 1997; Thompson, 1998a ; Thompson, 1998b ; Schmedgall & DeFranco, 1999; Sill & Decker, 1999; Thompson, 1999a ; Thompson, 1999b ; Kay, 2000)

그러나 앞선 문헌 연구에서 소개한 연구들은 거의 한국이 아닌 외국에서 진행된 연구로, 현재 까지 한국 내에서 집단 급식소를 대상으로 한 연구로는 급식으로 제공한 음식에 대한 영양가를 분석하거나 급식 관련 서비스 만족도 등의 연구가 주를 이루고 있고 이러한 연구들이 고객의 입장에서 조사가 진행되어 왔으며, 또한 국내에서 집단 급식 산업의 식수 인원 수요 예측과 관련되어 진행된 연구는 드물었다고 할 수 있다. (김종애, 2010 ; 최순태 2007 ; 박정소, 2011 ; 임재영 2016 ; 이슬, 2010)

이와 관련한 이유로서는 여러 가지가 있겠지만 그동안 한국에서의 연구가 실데이터를 기반으로 연구가 진행되기에는 어려운 환경에서 그 이유를 찾을 수 있다고 할 수 있다.

그리고 한식의 식단 구성이 서구식과 상이하고 복잡한 점을 들 수 있다. 이로 인해 다중 회귀 모델링 등 전통적인 통계 기법을 통해 식단 메뉴를 모델링을 하는 데에 여러 가지 제약이 발생하였을 것이라고 판단되어진다.

또한 식수 인원예 영향을 미치는 변수가 많아짐에 따라 식수 예측 모델링에 많은 제약 요소가 발생하고, 보다 정확한 모델링을 위해서는 중요한 설명 변수에 대한 더 많은 관측치를 구해 진행을 하여야 하기 때문에 이러한 조건들을 구비하여 양질의 실데이터를 기반으로 모델링과 연구를 진행하는 데 많은 애로 사항이 있었을 것으로 판단된다.

국내에서는 2000년대 들어서서 집단 급식소 중 대학교 급식소의 과거 식수 데이터를 근거로 식수 예측에 영향을 미치는 여러 요인을 분석하고, 이를 토대로 대학교 급식소의 식수 예측을 위한 수요 예측 모델을 제시한 연구가 있다. (정라나 2001)

이 연구를 통해 대학교 집단 급식소의 식수 인원은 봄 여름 가을 겨울 이라는 시계열에

다른 유형을 보이고 있었고, 대학이라는 특수성으로 인해 월별 학기별로 일정한 패턴을 보이는 것을 보여줌으로써 다른 집단 급식소의 예측과 차이가 있다고 하였으며, 이는 일반적인 집단 급식소의 연구라고 하기 보다는 대학의 구내 식당이라는 특수성이 주로 작용하여 진행이 된 연구라고 할 수 있다.

또한 이 연구에서는 대학 급식소의 식수 예측 모델링을 하는 경우에 있어서 식수 예측 관련하여 고려하고 있는 요인과 식수 인원의 상관 분석을 실시하였으며, 식수 인원은 1주전, 2주전, 3주전, 4주전의 식사 인원, 요일, 대학교 급식소의 이용률, 메뉴에 대한 선호도 간에 선형적으로 유의한 관련성이 있는 것으로 나타났다. (정라나 2001)

이 연구도 대학교 집단 급식소의 과거의 데이터를 바탕으로 다중 선형 회귀 분석을 이용하여 개발하였으나 식수 예측의 정확도나 개발된 Model Evaluation에 대한 구체적인 언급은 없다는 점에는 식수 예측 모델링 관련하여 한계점이 있다고 할 수 있다.

연구 개발된 모델이 실질적으로 적용하는 데 있어서 얼마나 정확하고 대학교 집단 급식소의 식수 예측에 기여하고 있는지에 대해 알 수가 없어 이 부분에 아쉬운 점이 있었으며, 향후 식수 관련 수요 예측에 얼마나 기여하는 지에 대해서는 추가적인 연구가 필요하다.

이외에도 사업체의 집단 급식소 식수 예측에 대한 연구가 있었다.

이 식수 인원 예측 관련 연구는 1차적으로 집단 급식소의 식수 예측에 영향을 미치는 요인들을 비교 분석하여 식수 오차율을 줄일 수 있는 개선 방안을 도출하고자 하였으며, 빈도 분석, t-test, 교차분석, ANOVA, 상관분석, 회귀분석 등 다양한 통계 분석을 통해서 수요 예측 방법들 간의 예측력을 비교 하여 집단 급식 점포에서의 식수 예측 방법을 제안하는 것을 목적으로 한 연구가 있었다. (임혜진 2008).

이 연구에서는 사업체 급식소 90개의 4개월간 끼니별 식수 자료를 수집하여 분석을 진행하였으며, 식수 예측에 영향을 미치는 요인을 도출하기 위해 온라인 설문 조사를 통해 해당 점포 관리자를 대상으로 점포 운영 자료와 식수 오차율에 대한 인식 정도를 조사

하기도 하였다. (임혜진 2008).

그 결과 식수 예측의 목적으로는 잔식 감소를 통한 원가 절감을 가장 중요하게 생각하고 있었으며, 고객 불만 감소, 조리 작업 효율화의 순으로 나타났다고 주장하였다. (임혜진 2008)

이 연구에 따르면, 식수 예측을 담당하는 경우 다수의 사람이 같이 참여하여 결정하기 보다는 영양사 단독으로 수행하는 경우가 대부분이었고 운영자, 조리사, 예측 전문가 등 다른 이와 같이 식수 인원을 예측하는 경우는 소수에 불과 하였으며, 식수 예측 시 고려하는 요인으로는 제공하는 메뉴에 대한 선호도, 날씨, 전주 동일 요일의 식수, 고객 사은 행사, 운영 끼니, 이전 1달간 식수 자료 순으로 고려하였다.

또한 식수 인원 예측에 영향을 미치는 변수를 도출하기 위해 상관 분석을 실시하여 식수 오차율에 대한 식수 예측 변수들의 회귀식을 도출하였으며 수요 예측 모델들의 식수 정확도는 회귀 방정식 모델의 예측력이 좋은 것으로 조사되었다. (임혜진 2008).

그리고 직접적인 식수 예측은 아니지만 식수 오차율에 영향을 주는 주요 요인들을 분석하고 식수 오차율에 영향을 미치는 기상 요인을 파악 하여 집단 급식 산업에서의 식수 예측력을 높이는 기여하고자 한 연구도 있었다. (임재영, 2015)

그러나 이와 같이 집단 급식소의 식수 인원 예측 관련하여 여러 가지 노력을 경주하여 왔으나 (김혜영 2000 ; 정라나 2001 ; 임혜진 2008 ; 임재영 2015), 아직 까지 식수 인원 예측을 위해 기계 학습 기법을 사용하여 집단 급식소의 식수 예측 모델링을 개발 연구한 것은 없는 것으로 조사 되었다.

본 연구에서는 식수 예측 모델링하는 데 있어서, 식수 예상 인원을 종속 변수로, 그리고 식수 수요에 영향에 미치는 여러 요인들을 독립 변수로 지정하여 이들 변수들 간의 여러 가지 모형으로 파악하고 기계 학습 기법을 이용하여 오차율을 최소로 하는 식수 예측 알고리즘을 개발하여 현장에서 도움이 될 수 있도록 한다.

집단 급식소에서의 데이터 사이언스 기반의 체계적인 식수 예측은 급식 서비스의 질을 높이는 선순환의 시발점이면서 앞선 여러 문제점을 선제 대응할 수 있도록 하는 일종의 네비게이션 같은 역할을 제공한다고 할 수 있다.



제 4 절 음식물 쓰레기 처리의 문제점

사회적 경제적 발달로 인해 현대인의 식생활 양식이 크게 변화하였으며, 특히 여성의 경제적 활동 증가와 맞벌이 부부의 증대로 인하여 집에서 식사 보다는 집 밖에서의 식사를 하는 경우가 많아지고 있으며 출산 인구의 감소와 여성의 사회적 진출이 점점 많아짐에 따라 조직 및 단체 생활이 늘어나면서 점점 집단 급식소가 점점 많아지고 있다. 이러한 추세에 따라 환경오염 문제가 미세 먼지와 함께 요즘 새로운 사회 문제로 대두되고 있다.

또한 먹는 것이 풍족해지고 식생활이 서구화 되면서 음식물의 과소비가 더욱 만연되고 있으며, 인간 수명이 연장됨과 동시에 결혼, 이혼, 1인 가구의 증가, 여자의 경제력 창출 등으로 인해 사회적 변화가 나타나고 이로 인해 식생활 소비에서도 많은 변화가 생겨났을 뿐만 아니라 이로 인해 사회적으로 많은 문제를 야기하고 있다,

특히 환경 오염 측면에서 일반적인 가정 생활과 집단 급식소에서 배출되는 음식 쓰레기 문제가 심각한 사항으로 인식되어 일반 가정 식사나 집단 급식에 있어서 음식 쓰레기 발생은 막대한 처리 비용이 발생하고 있어 비용 손실 뿐 만 아니라 사회적 환경 문제와도 바로 연결되기 때문에 가정에서 뿐만 아니라 집단 급식소에서 발생하는 음식물 쓰레기 처리에 대한 문제점이 지속적으로 지적되어 왔다. (김혜영 2000)

따라서 음식물 관련 쓰레기를 줄이는 것은 음식물 쓰레기를 처리하는 비용을 줄이는 직접 효과 뿐 아니라 간접적인 효과 측면에서도 과잉으로 제공으로 되는 음식물을 줄임으로써 부족한 식량 자원 절약에 이바지 하게 되고, 더 나아가 음식물 쓰레기로 부터 나오는 민원이나 주민간의 갈등을 줄일 수 있으며, 음식물 쓰레기로부터 나오는 여러 환경 오염원을 감소시켜 줌으로서 환경을 보호할 수 있다 점에서는 중요한 일이 되고 있다. (박세정 2017)

먹을 수 있는 식량 자원이 쓰레기로 버려지는 것은 두 가지 의미에서 문제가 되고 있다. 하나는 버려지는 자원의 낭비 부분이고 또 하나는 쓰레기 처리를 위한 또 다른 자원의 낭비가 발생한다는 점을 들 수 있다.

첫째, 우리나라와 같이 식량 자급률이 낮은 나라에서는 많은 식량 재료를 해외에서 수입하여 충당하는 실정인데, 음식물이 다 소비되거나 섭취되지 못하고 음식물 쓰레기로 버려진다는 의미는 그 만큼 수입을 줄일 수도 있는데 음식물 쓰레기 발생량 이상으로 수입을 하고 있다는 의미이다. 이는 직접적으로 귀중한 외화 낭비로 이어지는 것이고, 또한 버려지는 음식물 쓰레기가 국내에서 생산이 된 것이라고 하더라도 이는 음식물의 재료로 쓰여지기 위해서 사용되는 투입 분이 다른 곳으로 투입되어 사용될 수 있음에도 불구하고 잘못 사용되고 있다는 것을 의미하는 것이다.

둘째, 음식물이 섭취되지 아니하고 음식물 쓰레기로 버려진다는 것은 인건비가 상승함에 따라 음식물 쓰레기 수거 비용과 처리 비용이 증가하고 있어 사회적으로 여러 가지 문제를 야기한다. 또한 음식물 쓰레기를 매립지까지 수송하는 직접적인 수송 비용 등 직접 매립에 필요한 비용이 발생이 되며, 이로 인해 추가 비용이 발생하고 사회적으로도 많은 문제가 발생 등으로 매립의 확보가 어려워지고 있어 과거에 비해 현재에는 거의 이루어지지 않고 있다.

이렇게 음식물 쓰레기의 과다 발생으로 인해 식량 자원의 낭비와 쓰레기 처리 비용 증가로 인해 이중적인 경제적인 손실을 초래하고 있으며, 음식물 쓰레기는 직접적인 처리 비용과 손실 외에 음식물 쓰레기로 발생되는 추가적인 경제적, 사회적, 환경적 비용들을 총체적인 사회적인 비용으로 고려하여야 하며, 이를 종합적으로 고려하게 되면 그 비용은 매우 크다고 할 수 있다 (장종근 외 2001).

우리나라의 경우 음식물 쓰레기가 많이 발생하는 데는 다음과 같은 몇 가지 원인이 있다.

첫째, 우리나라의 상차림의 특수한 구조에 비롯하기도 한다. 벼농사를 근간으로 하여 전통적으로 내려오는 음식의 식습관 상 한국 음식은 다른 나라와 음식과 달리 기본적으로

주식·부식으로 구분이 되어 있으며, 또한 밥과 함께 국, 찌개, 그리고 여러 가지 밑반찬을 여러 가지로 차려 상차림을 하고 있다.

한국인의 서구화된 식습관으로 전통적인 한국의 상차림이 많이 줄었다고 하더라도 한국 전통적으로 내려오는 식생활 습관으로 인해 음식물 쓰레기가 발생될 수밖에 없는 상황이다. (송재성 1996 : 윤진숙 1996) 특히 음식점에서 제공하는 반찬의 가지 수와 양이 과다하여 제공되는 반찬과 식사를 다 섭취하지 아니하고 남게 되어 있어 식사 후에는 남은 반찬과 밥 등을 버리게 되므로 음식물의 낭비와 환경 오염을 유발하고 있는 것이 사실이다.

둘째, 한국 음식의 조리 방법에 있다. 서양 음식과 달리 한국 음식을 조리하는 과정이 정량화 혹은 계량화 되어 있지 않으며, 조리하는 데 있어 경험적으로 양을 정함으로 인해 쓰레기가 추가 발생하는 요인이 된다.

셋째, 일반적으로 우리나라 음식 문화에는 전통적으로 밥과 국을 동반하고 있어 우리나라 음식물 쓰레기는 수분 함량이 매우 높아 쉽게 부패가 될 뿐만 아니라, 이를 수거하고 운반하는 과정에서 발생하는 악취와 오수가 심할 뿐만 아니라 음식물 쓰레기를 매립 처리하는 경우에는 악취가 섞인 다량의 침출수를 발생시켜 지하수를 오염시키고 있고 (강효성 1994 ; 신명교 심흥균 ; 홍성철 1995 ; 이규만 1999), 이로 인해 2차 환경 오염을 유발하고 있어 현재 매립을 통한 음식물 쓰레기 처리는 거의 이루어지지 않고 있다.

넷째, 한국의 음식물 쓰레기는 물기가 많기 때문에 소각시에 보조 연료를 추가 사용하는 문제점을 가지고 있으며, 통상적으로 한국인의 섭식 습관상 음식물이 과다 염분 (3% 이상)을 포함하고 있는 음식물이 많아 이러한 과다 염분을 함유하는 음식물 쓰레기는 다른 음식물 쓰레기와 같이 섞여 버려진다고 하더라도 퇴비로 사용하기가 적합지 않아 이로 인해 한국 음식물의 쓰레기 처리 문제 있어서 비용은 지속적으로 증가하였다. (환경부 1997)

이렇게 음식물 쓰레기가 많이 발생할 수 밖에 없는 문제는 잘 해결되지 않고 있다. 일례로 음식물 쓰레기를 퇴비화하는 정책을 편다면 오히려 퇴비 발생량이 과다해지는 실정으로 이로 인해 음식물 쓰레기를 가지고 퇴비화하는 것은 현실적으로 많은 벽에 부딪히고 있어 음식물 쓰레기의 퇴비화에 대해 필요성을 느끼고 있으나 실질적으로 적용하기에는 매우 어려운 상황이 되고 있다. (이정임 외 1998 ; 환경 정보 서비스 1996)

또한 식당이나 가정에서 나오는 음식물 쓰레기를 퇴비로 사용하기 위해서는 음식물 쓰레기 그 자체의 상태로는 퇴비로 사용할 수 없어 이를 우선적으로 말려야 하며, 통상적으로 사용하는 고속 발효기 등의 처리 과정에서 나오는 냄새가 매우 심해 이는 민원과 지역 갈등 뿐 만 아니라 역한 냄새로 인한 2차 환경 오염을 야기시킬 수 있으며, 또한 이러한 발효기를 통해 음식물 쓰레기로부터 생성된 퇴비는 직접 토양에 사용할 수 있는 완숙 퇴비로 볼 수 없을 뿐만 아니라 고속 발효 되었다 하더라도 이 상태의 퇴비는 흙이나 토양에 뿌려지는 경우 해로운 점이 더 많이 나올 수 있다는 문제점을 가지고 있다. (전무영 외 2000)

그러므로 음식물 쓰레기는 단순히 음식물 처리 비용으로만 인식을 해서는 안되며 음식물 쓰레기 처리의 여러 가지 환경 비용 및 사회적 비용을 고려하여 좀 더 넓은 시각으로 음식물 쓰레기가 새로운 환경오염을 유발하는 새로운 원인으로 작용할 수도 있음을 인식하여야 한다.

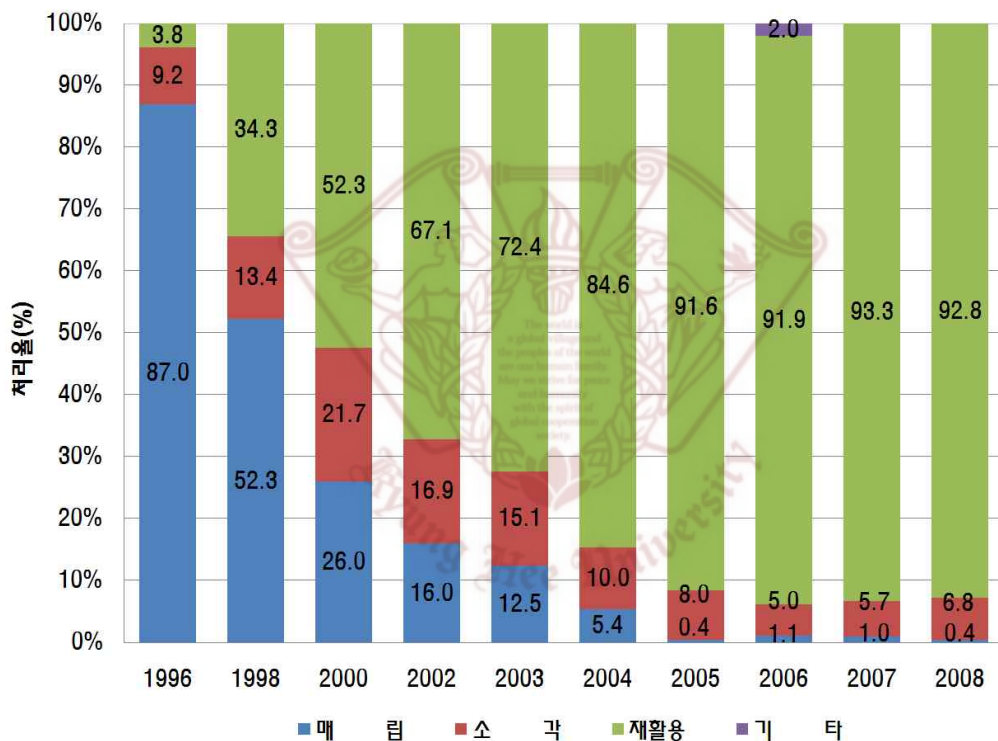
이러한 문제점으로 인해 음식물류 폐기물은 사회적 이슈로 주목받고 있는 바, 폐기물 관리법에 의해 2005년 부터는 음식물 쓰레기의 직매립 금지, 2006년 부터 음식물 쓰레기 처리 시설의 정기 검사 실시, 2013년부터는 음폐수 해양 매립 금지 등 전반적인 관리 규제가 강화되고 있다. (이정임, 2010, 경기 개발 연구원) 또한 수거 방식도 기존의 생활계 폐기물 혼합 수거 방식에서 분리 수거 방식으로 전환되었다.

그러나 음식물 쓰레기의 환경 친화적 처리는 또 다른 문제를 야기 시키고 있다. 최근 지역 NIMBY 현상의 극심화로 인해 음식물류 폐기물 처리 시설의 신규 설치가 매우 어려운 실정으로 서울시 및 수도권의 음식물류 폐기물이 자체적으로 처리되지 않고 오히려 인근

의 다른 시도로 민간 음식물류 폐기물 처리 시설로의 반입 처리가 증가하고 있는 상황으로, 타 시도로 부터의 음식물류 폐기물 반입이 늘어남에 따라 나타나는 폐수, 그리고 심한 악취 등 환경 오염 물질 배출로 인한 민원 발생과 환경 오염이 끊임 없이 증가하고 있는 추세이다. 서울의 음식물 쓰레기 일부가 경기도로 이관되어 처리되고 있다.

그림 2-1은 경기도 음식물류 폐기물의 처리 비율의 연도별 추이를 나타내는 그림이다.

그림 2-1 경기도 음식물류 폐기물의 처리비율 추이



(출처, 환경부 (2009). 『2008 전국 폐기물 발생 및 처리현황』. 이정임, 2010, 경기 개발 연구원)

이러한 음식물류 폐기물의 유입 경로, 처리자, 최종 처리 방법 등 반입 현황에 대한 보고 및 관리 체계가 미비한 가운데 폐기물의 해양 매립 금지에 따른 폐수 처리 비용 및 악취 처리 비용 등 음식물류 폐기물 처리 시설의 환경 처리 관련 예산이 지속적으로 증가되고

있는 실정이다. (이정임, 2010, 경기 개발 연구원)

음식물 쓰레기 발생량을 배출원으로 볼 때, 가정과 식품 접객업소가 가장 많은 편이며, 그 다음이 집단 급식소이어서 집단 급식소의 음식물 쓰레기 배출이 차지하는 비율은 전체 음식물 쓰레기 발생량의 배출원에서 볼 때 비교적 적은 편이라고 할 수 있다.

그러나 1인 가구의 증가, 맞벌이 부부의 증가 및 여성의 사회적 활동 그리고 전체 가구 인원의 대외 활동의 증가 등으로 인해 현대 사회에서 집단 급식소의 수는 계속적으로 증가할 수 밖에 없는 추세이다.

일반 접객업소나 일반 식당과는 달리 집단 급식소에서의 급식 대상은 기본적으로 그 인원이 내부 인원으로 운영되므로 사용 인원이 국한되어 있어 다른 음식 서비스 업체와 달리 관리하기가 쉽다는 장점이 있다.

그러므로 집단 급식소를 대상으로 잔반량을 줄이는 노력을 하는 것은 다른 음식물 쓰레기 배출원 중에서 가장 체계적이고 효과적으로 잔반량 줄이기 운동을 실시하기에 적합한 대상이라고 할 수 있으며, 집단 급식소에서의 잔반량 줄이기 운동을 통해 진행되는 여러 가지 방안은 향후 가정과 식품 접객업소에서의 잔반량 줄이기에 파급 효과가 클 것으로 생각되어 진다. (전무영 외 2000)

음식물류 폐기물 다량 배출 사업장은 폐기물 관리법 시행령 제8조 4항에 따르면, 영업장 면적 200m² 이상 혹은 평균 급식 인원 100인 이상 집단 급식소는 음식물류 폐기물 감량 대상 사업장으로 규정되어 음식물 쓰레기를 감량 배출하도록 되어 있다.

이처럼 법적으로 제도적으로 음식물 쓰레기를 줄이는 노력은 진행이 되고 있으나 일반 시민들이나 많은 사람들이 이러한 법이 있다는 사실도 모르고 있으며, 집단 급식소에서 이 시행령의 취지가 살아나기 위해서는 실효적인 측면에 있어서는 자발적인 참여 등 실질적인 참여가 중요하다

집단 급식소에서의 음식물 쓰레기 감량 방안으로는 단계별로 여러 단계에서 효율적으로 관리 할 수 있다.

집단 급식소에서의 조리 단계는 식단을 계획하는 단계, 식품 발주 단계, 식품 구매 단계, 식품 검수 단계, 음식물 전처리 단계, 음식물 조리 단계, 배식 단계, 퇴식 단계, 잔식 처리 및 잔식 재사용 단계 그리고 잔반 등 음식물 쓰레기 배출 단계 등으로 이루어진다.

이는 식단을 계획하는 단계부터 음식물 쓰레기 배출 단계의 전 단계 이르기 까지 다양한 방법에 의해 음식물 쓰레기 감량 방안이 이행될 수 있는데 이를 표로 나타내면 표2-2와 같다

표 2-2 집단 급식소의 급식 생산 단계별 음식물 쓰레기 감량 방안

단계	내용
계획 단계	대상자의 만족과 영양 권장량에 맞게 식단 작성을 하는 것이 필요
발주 단계	식수 인원 수요 예측 단계로 현장에서 필요한 식수 인원 파악과 표준화된 음식 요리법을 활용, 현장에 필요한 식자재 산출이 중요
구매 단계	식자재의 신선도가 좋으면서 음식물 쓰레기가 상대적으로 적게 발생하는 상태의 식재료나 포장 제품을 구매
검수 단계	제공되는 식재료에 대한 품질과 양에 대한 정확한 검사를 실시하며 구매하고자 하였던 식자재의 내용이 식자재 규격서와 일치하는지를 확인하고 또한 식자재의 신선도가 유지되도록 하는 것이 중요
전처리 단계	제공되는 식자재 관련한 위생 상태와 신선도를 고려하여 제공할 음식 서비스를 위한 식자재 전처리 실시
조리 단계	집단 급식소의 현장 특성에 맞는 표준화된 레시피 및 조리법을 적용
배식 단계	고객 스스로 할 수 있는 자율 배식을 하거나 또는 특정 음식에 대한 부분 자율 배식으로 잔반량을 줄임
퇴식 단계	식사후 퇴식구에 잔반 줄이기 운동 관련한 지도 및 홍보를 하고, 제공 되는 음식 제공 서비스에 대한 일일 잔반량을 매번 체크하여 다음 급식 서비스를 위한 생산 계획에 참고하여 진행

잔식 처리 및 재사용 단계	통상적으로 남는 잔식은 버리지 말고, 어려운 사람들을 위해 음식을 제공하는 기부 단체에 기탁함과 동시에 신선도가 유지되는 수준에서 재사용이나 재활용이 가능한 남은 음식물은 위생적으로 관리하여 다음 식사 서비스 제동 때 재활용
음식물 쓰레기 배출 단계	잔반 관련한 수분을 줄이고, 잔반 외의 비닐, 종이, 이쑤시개, 은박지 등 이물질들을 제거하여 자원화에 지장을 주지 않도록 하며, 또한 음식물 쓰레기 감량 의무 사업장은 음식물 쓰레기를 감량화 처리 후 배출

음식물 쓰레기를 줄이는 노력은 환경 의식의 변화를 통해 줄이는 방법이 음식물 쓰레기의 발생을 원천적으로 줄일 수 있는 가장 좋은 방법 중 하나가 될 수 있으며, 이는 시간이 오래 걸리고 또한 이를 전개하는데 있어서 많은 홍보 비용이 든다는 단점이 있지만 지속적으로 실시를 해야 한다.

음식물 쓰레기를 줄이는 또 다른 방법은 집단 급식소의 식수 예측을 통해 음식물 쓰레기 자체를 원천적으로 줄여주는 방법으로 이러한 방법은 집단 급식소의 식수 인원 예측을 통한 방법을 새로운 대안으로 제시할 수 있다.

이는 음식물 쓰레기 발생을 줄이는 선제 대응 일환으로 실무적으로 적용이 되어 많은 파급 효과가 있을 것으로 판단된다.

식수 예측 모델링을 통해 식수 예상 인원에 맞추어서 식재료를 준비하여 당일에 맞는 음식을 조리하는 방법이 현실적이고 직접적이고 많은 사람의 동참을 끌어낼 수 있는 좋은 방법이 될 것이라고 판단된다.

또한 식수 예측 기법은 시간이 지남에 따라 더욱 정교해짐으로써 일정 시간이 지나면 단순한 식수 예측 뿐 만이 아니라 영양 추천 서비스 등 다양한 서비스로의 확장도 가능해지는 장점이 있다

이상 앞서 살펴 본 바와 같이 음식물 쓰레기를 줄이기 위해 식수 예측을 통한 선제 대응 방식과 음식물 쓰레기 처리 사후 방식 진행을 동시에 함으로써 앞서 기술한 음식물 쓰레기의 문제점 및 환경 오염, 사회적 비용 등 여러 제반 비용을 고려해 볼 때 음식물 쓰레기를 줄이는 노력은 더욱 더 절대적으로 필요로 한 것이며, 음식물 쓰레기로 인해 발생하는 음식물 처리 비용의 증가, 악취 등으로 인한 민원, 그리고 처리 과정에서 발생하는 환경 오염 등 환경의 변화와 비용을 줄이는 노력과 의식을 가지는 것은 매우 중요하다고 할 수 있다.



제 5 절 집단 급식소의 음식물 쓰레기

집단 급식소에서의 급식 대상은 기본적으로 그 인원이 내부 인원으로 운영되므로 다른 음식 서비스 업체와 달리 관리하기가 쉬우며, 집단 급식소를 대상으로 잔반량을 줄이는 노력을 하는 것은 다른 음식물 쓰레기 배출원 중에서도 가장 체계적이고 효과적으로 잔반량 줄이기 운동을 실시하기에 적합한 대상이라고 할 수 있다. (전무영, 민혜성 2000)

집단 급식소에서 내부 인원, 직원 등 고객에게 식사를 제공하는 데 있어서 음식물 쓰레기는 크게 3가지로 분류가 된다.

첫째, 처음 입고된 식품을 조리하기 위해 처리할 때 발생하는 음식 쓰레기인 음식 전처리 쓰레기,

둘째, 배식시 잘못된 1인당 1회 분량이나 고객의 기호이나 취향에 맞지 않은 음식으로 고객이 먹고 남긴 것인 잔반 (plate waste),

셋째, 그리고 식수 수요 예측의 부정확 등으로 야기된 초과 생산으로 고객에게 다 제공되지 못하고 남은 음식인 미배식 잔반인 즉 잔식 (overproduct)으로 구성되며,

전처리 쓰레기, 먹고 남긴 잔반, 그리고 미배식 잔반인 잔식을 통틀어서 광의의 개념으로 잔반으로 이라고 한다.

집단 급식소는 조리 기획 단계부터 식사 후 음식물 쓰레기로 배출되기 까지 음식물 쓰레기의 발생은 필연적으로 나타날 수 밖에 없으며, 많은 집단 급식소에 있어서 이러한 급식 준비를 위한 전처리 쓰레기, 먹고 남긴 잔반, 그리고 미배식 잔반인 잔식 등 음식물 쓰레기를 줄이는 노력은 지속 되어 왔다.

특히 고객의 선호도에 맞는 메뉴 개발, 잔반을 줄이자는 많은 홍보 활동을 통해 식당을

이용하는 고객에게 잔반 제로 운동 등을 펼치는 등 통해 많은 노력을 하고 있으며, 급식 준비하는 데 나오는 전처리 쓰레기를 실질적으로 줄이기 위한 노력을 하여 왔다.

그러나 대다수의 집단 급식소에서의 미배식 잔반인 잔식은 식수 예측이 잘못됨에 나타나는 음식물 쓰레기로써 이는 영양사, 조리사 등의 경험치에 의존하여 진행을 하고 있어, 이를 통제하는 것은 매우 어려운 상황으로 인식되고 있어, 잔식은 불가피하게 발생하는 음식물 쓰레기 비용으로 관행적으로 인식하여 처리를 하고 있다.

통상적으로 음식물 쓰레기는 전처리 쓰레기, 먹고 남긴 잔반, 그리고 미배식 잔반인 잔식 등의 음식물 쓰레기를 처리하는 비용으로만 인식하면 안되며, 이러한 음식물 쓰레기로 인한 직접적인 처리 비용 뿐만 아니라, 과도한 재료 구입비, 추가되는 노동력 등의 직접 비용으로 인식하여야 하며, 더 나아가 이로 인해 발생하는 환경 오염 등 사회적 비용을 감안하면 더 더욱 그 비용이 커질 수 밖에 없다.

본 연구의 대상인 S시청 구내 식당의 잔반 또한 구성하는 요인은 크게 조리시 전처리 잔반, 식사 후 남은 잔반, 그리고 배식 후 남은 미배식 잔반인 잔식 등 크게 3가지로 나뉘는데, 이때 일일 잔반량에 가장 큰 영향을 끼치는 것은 전체 잔반량의 50%를 차지할 정도의 많은 비중을 차지하는 미배식 잔반인 잔식이라고 할 수 있다.

급식소에서 발생하는 쓰레기 중 미배식 잔반인 잔식을 줄이는 것은 급식 서비스 선순환 시스템의 중요한 요소 중 하나이며, 정해져 있는 판매가와 제한된 식품 재료비로 운영되는 집단 급식소의 특성 상 잔식 발생 및 처리에 드는 직간접적인 재정적인 손실과 기회 비용은 관리해야 할 중요한 요소이며 이는 간과할 수 없는 실정이라고 할 수 있다. (김혜영 2000)

S 시청의 집단 급식소인 구내 식당은 직원들의 식사를 위해 조리를 하기 위해 발생하는 음식 전처리 쓰레기와 식사 후 남은 잔반은 집단 급식소인 S시 구내 식당의 자체적인

노력과 직원들의 협조를 통해 조정 가능하며, 일일 잔반량에 미치는 영향도 작은 것으로 파악 되었다.

그러나 예상하지 못한 식사 인원으로 부터 발생하는 미배식으로 인해 남은 잔반인 잔식은 인원이 식당에 올 것으로 예상한 것과 다르게 방문했을 때 생기는 잔식으로 평상시에 예측한 대로 실무자들이 처리 할 수 있는 허용 범위 내에서 식사 인원이 오다가 갑자기 많이 오는 예상 식수 인원의 예측 불균형으로 발생하고 있다.

이러한 경우가 발생 시 예측 하지 못한 일들을 처리하기 위해 취해지는 여러 조치는 현장에 계신 분들에게 많은 스트레스를 야기시키고 또한 민원으로 이어진 부분에 대해서 이를 해결하기 위해 많은 노력이 추가적으로 필요하다고 할 수 있다

집단 급식소인 S시 구내 식당의 잔반을 줄이기 위해서는 평상시에도 조리를 위한 기획 단계에서 부터 적합한 식수 예측이 진행이 되어야 하며, 이를 위해 경험적으로 진행되어 온 식수 예측을 바탕으로 이를 데이터 과학의 체계적인 분석 및 예측을 함으로써 적합한 수요 예측을 근거로 음식물 쓰레기인 잔반을 줄이고 식사의 질을 높이는 지속적인 노력이 필요한 상황이라고 할 수 있다.

제 6 절 메뉴의 중요성과 메뉴 분류 체계

1. 메뉴의 개념과 메뉴 분석의 중요성

메뉴는 식사의 요리 종류를 나타내는 것으로 사전적 의미로는 식사로써 제공되는 음식들에 관해 상세히 기록한 표 또는 음식점 등에서 파는 음식의 이름과 값의 표를 의미한다 (나정기 2009).

메뉴는 고객에게 식사로 제공되는 요리의 품목, 명칭, 형태, 순서 등을 알기 쉽게 설명해 놓은 목록이나 차림표를 말하는 것으로, 한국어로는 ‘차림표’ 혹은 ‘식단’ 이라고 한다. (오정환 1984)

이러한 메뉴는 시대에 따라 그 의미가 항상 변화되어 왔으며, 1960년대에는 메뉴를 단순히 차림표로 정의되었고, 1970년대에 들어와서는 메뉴를 마케팅의 개념이 포함된 차림표로 정의되었으며, 1980년 이후 부터는 차림표의 개념은 약해지고, 메뉴를 강력한 마케팅 및 내부 통제의 톨로 정의하였다. (나정기 2009)

실질적으로 고객들이 음식점을 선택할 경우 음식 맛을 보고 선택하는 경우도 있지만 실질적으로 대다수의 고객들은 메뉴를 보고 선택을 한다고 할 수 있다. 그러므로 이 경우 고객에게 제공되는 메뉴는 고객에게 서비스를 사용 할지 말지를 결정하도록 해주는 역할을 하는 것이라고 할 수 있다.

그러므로 음식 서비스 산업에서 메뉴는 단순한 경험의 산물로 여겨 메뉴를 구성을 하는 것에서 탈피하여 메뉴를 식당으로 운영하는 데 있어서 다양한 분석과 운영의 결과의 노하우의 산물로 만들어 가는 것이 중요하다고 할 수 있다.

집단 급식에서도 메뉴는 가장 중요한 판매 도구이자 주요 통제 요소로 제공되고 있다고 할 수 있다.

집단 급식소에서 메뉴는 과학적인 메뉴 분석을 통해 고객의 메뉴 선호도와 욕구를 파악하는 식당 운영 시스템의 핵심이 되는 기능이며, 메뉴의 과학적이고 효율적인 관리와 고객에 맞는 메뉴 제공은 필수적인 사항이라고 할 수 있다. (Spears, 2000)

또한 메뉴는 고객의 입장에서 볼 때 식사 여부를 할지 말지를 결정하는 아주 중요한 것으로 소비자들이 받는 이미지, 느낌, 메뉴에 표시된 가격, 예상되는 품질 및 서비스는 이것을 보고 구매하려는 고객에게는 메뉴가 제공되는 음식 서비스를 통해 기대하는 가치로 여겨진다고 할 수 있다. (최현정 2010)

그러므로 메뉴는 식사 서비스를 제공하는 식당에서 상품 자체의 설명과 가치 증진을 위하여 필요한 것이며, 고객의 만족을 증대 시켜 주기 위해서 제공되는 품목과 형태를 체계적으로 짜놓은 차림표를 의미한다. (원윙희 1989)

또한 메뉴는 고객에 대한 식당의 모든 약속을 집합 시켜 놓은 하나의 표이며, 판매, 광고, 판매 촉진을 포함하는 마케팅의 도구라고 할 수 있다. (나정기 1998)

앞서 보듯이 국내에서도 메뉴에 대한 많은 정의가 있어 왔으며, 해외에서도 많은 학자들 사이에서 메뉴에 대한 정의가 있어 왔다.

메뉴는 식당 운영을 함에 있어서 기본적인면서도 아주 중요하면서 중추적인 역할을 하는 도구라고 정의를 하였다. (Mahmood A. Khah, 1989 ; Judy L. Miller, 1992 ; Jendal H. Kotschevar, 1975 ; Jack E. Miller, 1992 ; John W. Stokes, 1982 ; Edward A. Kazarain, 1989 ; Jack D. Ninemeier, 1986)

또한 메뉴를 단순한 차림표가 아닌 식당의 판매, 홍보 및 PR 도구 이면서 대화의 채널이자 식당 운영에 있어서 중요한 마케팅의 도구로 인식하고 정의를 하였다. (Dougals C.

Keister, 1977 ; William L. Kahrl, 1978 ; Bernard Davis & Sally Ston, 1991 ; Konald B. Lundberg, 1986 ; David V. Pavesic, 1989 ; Nancy Noman Scanlon, 1990 ; Judi Radice, 1987 ; Albin G. Seaberg, 1991)

또한 메뉴가 가지는 정보 제공의 기능과 가치를 중요시하여 메뉴는 정보의 제공자라고 정의 하였으며 (Haryr Berberoglu, 1987), 식당 운영에 있어서 가장 중요한 고객의 중요성을 인지하고 이를 중시하여 메뉴를 레스토랑과 고객을 연결하는 대화의 고리이자 그리고 메뉴를 커뮤니케이션 도구라고 정의를 하였다. (Lothar A. kerch, 1984 ; Leonard F. Fellman, 1981)

해외에서 학자 별로 메뉴를 정의한 것에 대해 정리한 한 것은 표 2-3과 같다

표 2-3 메뉴의 정의

학자	기본 정의
Judy L. Miller (1992) John W. Stokes (1982) Edward A. Kazarain (1989) Jack E. Miller (1992) Jendal H. Kotschevar(1975) Mahmood A. Khah (1989) Jack D. Ninemeier (1986)	메뉴는 식당이나 레스토랑 등 음식 서비스를 제공하고 운영함에 있어서 매우 핵심적인 역할과 운영상의 중추적인 역할을 담당하는 관리 도구이면서 또 다른 통제 도구 이다.
Dougals C. Keister (1977) William L. Kahrl (1978) Konald B. Lundberg (1986) Bernard Davis & Sally Ston(1991) Judi Radice (1987)	메뉴는 제공되는 음식 서비스에 대한 판매 도구 이다
Anthony M. Rey et all (1985) Robert A. Brymer (1987)	메뉴는 가장 중요한 마케팅 도구이다.

Haryr Berberoglu (1987)	메뉴는 음식 서비스 등 여러 가지 정보의 제 공자이다
Leonard F. Fellman (1981) Lothar A. kerch (1984)	메뉴는 음식 서비스를 제공하는 레스토랑과 식당에서 고객과의 연결을 이어주는 대화의 고리이자 메뉴는 소통 및 커뮤니케이션 도구 이다
Nancy Noman Scanlon (1990) David V. Pavesic (1989) Albin G. Seaberg (1991)	메뉴는 레스토랑이나 식당의 고객과의 대화, 제공되는 음식 서비스의 판매 그리고 홍보 및 PR 도구이며, 내부의 중요한 마케팅 도 구이다.

출처 : 김기영 외 2인, 식음료 서비스 실무론, 대왕사 1999 p.109.

최현정 (2010), 메뉴 엔지니어링 및 판매량 추이 분석을 통한 메뉴 운영 전략 개발,
나정기 (1994), 메뉴 계획과 디자인의 평가에 관한 연구, 경기 대학 대학원 박사 논문,
재인용 및 재구성

이러한 메뉴의 중요성을 기본으로 하여 메뉴에 대한 분석 및 메뉴 엔지니어링과 같은 메
뉴 분석 기법에 여러 연구가 있어 왔으며, 실무적으로 메뉴 엔지니어링 기법을 사용하여
왔다.

메뉴 분석은 제공 되고 있는 음식 서비스 관련하여 현 메뉴의 효율성을 평가하고, 메뉴의
가격 및 그 내용과 디자인을 결정하는 도구로서 고객 선호도와 내부 관리 능력을 평가하
여 매출량을 증대시키고 순이익을 증가시킬 수 있는 메뉴 아이템들의 평가라고 할 수 있
다. (진양호 1998)

그러므로 단순한 메뉴 분석에서 더 나아가 메뉴 엔지니어링이란 앞으로 나갈 식음료의 사
업 방향을 결정하기 위해 정보를 획득하고 메뉴의 구성, 모양, 식사 제공의 수익성이나
대중성 등의 측면에서 적정성을 평가하고 판단하는 것과 관련된 활동으로 정의가 되고,
메뉴 분석이 되지 않은 메뉴 개발은 메뉴로서 거의 의미가 없을 정도로 메뉴 분석이 매
우 중요하다고 할 수 있다. (이정자 1999)

문헌 연구를 통해 살펴 볼 때, 메뉴의 분석을 실무적으로는 부분적으로 사용하여 왔으나, 실질적으로 식수 예측에 전면적으로 적용한 사례는 적었다고 할 수 있다.

메뉴 분석을 기반으로 하여 고객의 메뉴 선호도를 실질적으로 모델링에 반영하는 것은 식수 예측에 있어서 중요한 요소가 되고 있다고 할 수 있다.

본 연구에서는 제공되는 메뉴의 분석을 토대로 하여 메뉴 특성 변수가 식수 인원 예측을 위한 전체 모델링 관련한 설명력을 파악하고 식수 예측에 영향을 미치는 다른 변수들과 종합하여 실제 메뉴의 실데이터를 기반으로 한 식수 예측 모델링을 진행하고자 한다.



2. 통일되지 않은 한식 메뉴 분류 체계

한식이란 한국 음식의 준말로 한국에서 전통적으로 사용되어 온 식재료 및 그와 유사한 식재료를 사용하여 한국 고유의 조리 방법 또는 그와 유사한 방법으로 만들어진 음식으로 한국 민족의 역사적, 문화적 특성을 가지고 생활 여건에 알맞게 창안되어 발전, 계승되어 온 음식을 뜻한다. (이민아, 2008)

세계 여러 나라의 식생활을 살펴 보면 어떤 한 민족의 식생활의 문화는 그 민족이 거주한 지역의 날씨, 문화 그리고 정치 경제 교육적 배경에 따라 형성되고 발전해 왔다고 할 수 있다.

특히 우리 나라는 사계절의 변화가 뚜렷하고 3면이 바다로 둘러 쌓여 있어 해산물이 풍부하고 젓갈류, 장류, 김치류 등 발효 식품 및 저장 기술의 발달로 인해 계절과 지역에 따른 특성에 따라 한국만의 독특함이 있고, 그 방법도 다양해서 한식을 어떻게 분류하느냐가 매우 중요하다.

한식의 분류가 중요함에도 불구하고 한식의 세계화가 진행되었던 정부에서도 한식을 분류하는 것에 대해 여러 연구를 하여 왔지만 한식 메뉴의 분류 체계에 대해서는 아직도 복잡하고 학자와 기관에 따라 사용되는 분류 체계가 달라 공통적으로 통일되게 사용되고 있는 방식이 없다고 할 수 있다.

이에 한식의 역사를 살펴 볼 때, 우리나라 한식 문화 역사 연구에서 식생활에 관한 연구는 한반도의 긴 역사에 비하여 기록으로 남겨진 것이 매우 드물며, 또한 기록으로 남겨진 문헌들도 상당 부분이 중국 문헌들을 참고하여 쓰여진 조리서이고, 우리말로 된 한글로 쓰여진 조리서 (調理書)는 그 수가 많지가 않다.

우리나라에서는 조선 시대에 와서야 다채로운 고조리서를 접할 수 있게 되며, 삼국 시대나 고려 시대의 고조리서는 거의 찾아 볼 수 없는 상황이다.

지금 까지는 조선 시대 어의를 지낸 전순의가 1450년 경 지은 「산가요록, 山家要錄」이 음식 관련 조리법을 기록한 책으로써 현재 까지 발견된 가장 오래된 책으로 간주하고 있으며, 이후 조선 중기 중종 때인 유학자 1540년경 안동의 김유(金綏 1491-1555)에 의해 집필된 「수운잡방(需雲雜方)」이 있으며, 1600년대에 들어서서 조선 시대의 조리서로는 허균(許筠) (1569-1618)이 광해군 시대의 팔도 지역 음식을 기록한 「도문대작 (屠門大嚼)」, 조선 중기의 문신 학자인 이수광(李光)이 지은 「지봉유설(芝峰類說)」(1613년), 사대부가의 음식을 직접 담당했던 장계향이 쓴 「음식디미방」(1670년경)이 있다. (정혜경 외 2015)

16세기에 저술된 「수운잡방(需雲雜方)」과 17세기에 기술된 「음식디미방」은 약 100년간의 시간적인 차이를 두고 쓰여진 책으로, 「음식디미방」과 「수운잡방(需雲雜方)」에 수록된 음식의 옛날 조선 시대의 조리법의 비교를 위해서 음식을 주식과 부식, 떡 및 한과, 주류, 발효 식품, 음청류, 조미료 그리고 기타 등으로 분류하여 비교를 한 연구가 있었다. (정혜경 외 2015)

특히 부식을 일반적인 분류 기준에 따라 탕이나 국, 그리고 찌개·지짐이, 채 혹은 나물, 생채나 무침, 누르미, 찜·조림·초, 볶음이나 구이, 적·선·전·포 그리고 젓갈·편·회, 침채·지히, 자반 등으로 분류하였다. (정혜경 외 2015)

한문이 아닌 한글로 쓰여진 조리서로서 「음식법 (饌法)」라는 책이 있는 데 이 책이 쓰여진 시기는 조선 시대 철종 5년으로 추정하는 데 이 책은 조선의 어느 가정에서 필사를 해서 혼인을 하는 딸에게 주는 요리서이다,

「음식법 (饌法)」에 나오는 모든 음식은 주식류, 찬물류, 떡류, 과점류, 음청류, 술류, 양면류 등으로 분류를 하였으며, 여기에서 사용된 음식명은 대체적으로 조리 방법에 따라 분류를 했다. (박미자, 1996)

주식류는 밥, 면류, 그리고 만두류로 분류를 하였으며, 밥은 약밥 1종류 6가지 재료로, 면

류는 냉면, 난면, 약면, 흰떡국수 등 4가지로 그리고 만두류는 전복 만두, 건치 만두, 어 만두, 제육 만두, 수교아 등 5가지 종류로 구분하였고 주식류의 종류, 재료, 빈도를 표로 나타내면 표 2-4와 같다.

표 2-4 「음식법 (饌法)」에 나오는 주식류의 종류, 재료, 빈도

분류 (종류)	음식명의 종류	재료수 (가지)	재료와 빈도
밥(1)	약밥	6	잣, 대추, 찹쌀, 꿀, 참기름, 간장
면류 (4)	냉면	12	무, 배추, 상육(등심), 배, 밤, 미나리, 파, 고추, 유차, 석류, 계란(2), 잣
	난면	11	후춧가루, 밀가루, 오이, 잣, 계란(2), 닭백숙, 오미사국(2), 표고, 송이, 깨국(2), 생선지짐
	약면	11	꿀, 기름, 후추 가루(2), 잣(2), 녹말, 생콩가루, 숙제육, 토장 깨국이나 오미자 국, 장국, 계란
	흰떡국수	2	파, 흰떡국수
만두류 (5)	전복만두	4	문어, 잣, 전복(2), 추포
	건치만두	6	후춧가루, 생치, 꿀, 기름, 잣, 간장(3)
	어만두	19	후추, 송어, 잣, 두부, 기름장, 소금, 간, 양, 국화잎, 제육(2), 천엽, 생강, 실고추, 파, 녹말, 표고, 전복(2), 송이, 생대하
	제육만두	4	잣, 제육 안심, 기름, 간장
	수교와	11	초장(황육, 영계, 생치밀 등 고음국 사용), 가루, 후추, 소금, 잣, 기름(3), 오이, 닭, 석이, 고기, 표고(2)

출처 : 박미자 (1996) “음식법 (찬법)”의 조리학적 고찰

「음식법 (饌法)」의 찬물류는 반찬 등을 나타내는 것으로 찬물류에는 국이나 탕, 찜, 전골, 선, 편육, 지짐이, 무침, 자반 및 부각, 전약 및 족편, 느르미, 채, 김치 등이 있다. (박미자, 1996)

탕은 소잡탕, 잡탕, 깨국잡탕 등 3가지로, 전골은 열구자탕 1종류로, 찜은 생치찜, 개장, 송이찜, 생복찜, 대하찜, 표고찜, 다시마찜, 묵찜, 게찜 등 9종류로 분류를 하였다

초는 해포(게포육) 등 3가지로, 무침은 대하 등 1종류로, 전약 및 족편류는 전약 및 족편 2종류로, 편육은 제편 1종류로, 자반 및 부각은 섭좌반, 임자좌반, 각종 부각 등 3종류로 지짐이는 생선 지짐 등 1종류로 분류를 하였다

느르미는 생치, 제육 느르미, 느르미, 동화 느르미, 낙지 느르미, 생선과 달걀 느르미, 게 느르미 등 5가지로, 선은 무우선과 배추선 등 2종류로, 채는 소채, 화채 등 2가지로 김치는 침채 등 1가지로 음식의 종류를 분류하였다.

떡류는 찐떡, 찰떡, 지진떡 등 3분류 하였고, 찐 떡은 시루떡, 당귀떡, 메꿀떡, 석이떡, 소꿀찰떡, 두텁떡, 송편, 증편 9종류로 구분을 하였고, 쑥단자, 석이단자, 당귀단자, 토란단자, 준수단자, 마단자, 수란떡, 잡과편 등 8가지로 그리고 지진떡은 밤조약, 당귀조약, 대추조약, 더덕산승, 생강산승 등 5가지 종류로 구분을 하였다.

과정류는 총 6가지로 구분하여 엿으로 만든 강정, 유밀과, 다식, 유과, 전과, 과편 등으로 분류를 하였고 음청류는 3가지로 구분하여 수정과 화채, 밀수 등 분류를 하였으며, 양념은 4가지로 구분하여 꿀, 장, 깨, 기름 등으로 분류하였고, 술은 도인주, 국화주, 송앵주, 인동주 등 4종류로 분류를 하였고, 조리법에서 사용된 용어를 보면, 조리 기법을 3가지로 구분하여 준비하는 과정과 써는 과정 그리고 음식을 익히는 과정 등으로 분류하여 사용하고 있었다. (박미자 1996)

위의 예에서 보듯이 식품 이름, 조리, 가공 용어 등이 시대와 지역에 따라 변한다는 것을 알 수 있으며, 현재 시점에서 보면 명칭과 실물을 결부시키기가 어려운 점이 많다고 할 수 있다.

다만 위의 예에서 음식의 분류 관련하여 추론할 수 있는 방법은 한식은 밥과 면 등 주식류와 국이나 탕 등의 국물류 그리고 여러 가지 조리 방법으로 사용하여 만들어진 찬류 그리고 후식이나 간식으로 먹는 떡류, 과정류, 음청류 그리고 기타 음식으로 볼 수 있는 양념, 주류 등으로 분류가 되어 있다는 점이다.

우리나라 한식의 조리상의 특징은 주식과 부식이 확연히 분리되어 발달하였다는 점이다. 이는 그동안 우리나라 한식에서는 곡물 조리법이 다양하게 발달을 하였으며, 이를 통해 음식의 종류가 많고 조리법이 아주 다양하다는 알 수 있다.

그리고 한식을 밥, 죽, 미음, 응이, 암죽, 죽, 국수류, 만두, 떡국 등의 주식류와 부식류로는 국, 탕, 찌개, 전골, 찜, 조림과 초, 볶음, 구이, 전, 산적, 편육, 족편, 순대, 나물, 쌈, 회, 숙회, 마른 반찬, 젓갈, 김치 등 부식류로 구분하여 음식을 분류하였고, 그 조리 방법에 따른 영양 및 음식의 가치를 연구하였다고 할 수 있다. (김세연, 2007)

또한 조화로운 맛을 중요시 여겨서 향신료와 조미료의 이용이 많으며 저장 식품이 발달하였고 모양을 중요시 하는 서양 요리와는 달리 맛을 위주로 하는 조리 문화가 발달되어 왔다고 할 수 있다. (노수정 2000)

또 다른 방법으로는 한국 음식을 그 형태 및 조리 방법에 따라 분류하고 있으며, 이는 밥류, 면 및 만두류, 떡 및 죽류, 국류, 찌개류, 찜류, 구이류, 전, 격, 부침류, 볶음류, 조림류, 튀김류, 나물 및 무침류, 김치류, 회류, 적갈류, 장아찌류, 장류, 과일류, 조리 않고 먹는 음식, 당류, 유지류, 다류, 건강 식품 및 양념류로 분류하기도 하였다. (한국 식품 산업 연구소, 1991)

또한 최근에 들어와서 메뉴의 선호도를 업종에 따라 한식, 중식, 일식, 양식, 뷔페, 약선 음식, 분식 등으로 분류하고, 그리고 식재료에 따라 쇠고기 요리, 돼지 고기 요리, 야채류 요리, 해물 요리 등으로 분류하고, 조리 방법에 따라 국물류, 볶음류, 튀김류, 찜류, 구이류 등으로 분류하고 음식의 맛에 따라 매운 음식, 담백한 음식, 짭짤한 음식, 얼큰한 음식 등으로 분류하고 음식의 온도에 따라 뜨거운 음식, 차가운 음식 등으로 분류하기도 하였다. (이원갑 외 2014)

좀 더 체계화되고 단순한 방법으로 분류 방병으로 한국 음식을 주식류, 부식류, 떡과 한과류, 음청류, 그리고 주류로 크게 5가지로 나누어 분류를 하고 있는 것도 있는 데 이를

좀 더 세부적으로 설명을 하면 다음과 같다. (노수정 2000)

주식류는 흰밥, 잡곡밥, 그리고 채소류와 어패류, 육류 등을 같이 넣어 만드는 별미밥, 비빔밥 등 밥이 있다.

또한 밥 이외에 죽, 이음, 응이, 암죽, 즈 등이 있으며, 응이는 곡물을 곱게 갈아 전분을 가라 앉혀서 가루로 말렸다가 물에 풀어 익혀 마실 수 있을 정도의 농도로 끓이는 것이다. 암죽은 모유가 부족할 때 아기를 키우던 대용식 및 이유식으로 쓰였다.

밥과 죽과 즈 외에 주식으로는 국수류가 있으며 국수류는 밀국수, 메밀 국수, 녹말 국수, 찹 국수, 강낭 국수, 썩 국수, 미역 국수 등이 있으며 또한 국물이 들어간 온면과 냉면이 있고 그리고 비벼 먹는 비빔면이 있다. 또 다른 주식으로는 만두, 떡국 등이 있다.

이 분류 방식에 따르면 주식은 밥류, 죽류, 국수류, 만두, 떡국 등이 있다고 할 수 있다.

부식류로는 국이나 탕, 찌개 및 전골, 선, 찜, 조림과 초, 볶음, 구이, 전, 산적, 편육, 족편, 순대, 나물, 생채, 쌈, 회, 숙회, 강회, 수란, 마른 반찬, 젓갈, 식혜, 김치 등이 있다. (김세연 2007)

떡류는 시루에 찌는 떡, 안반이나 절구에 치는 떡, 가루를 반죽하여 모양을 만들어 빚는 떡, 번철에 지지는 떡 등 곡물, 고물의 종류나 재료, 기본 조리법, 지역적인 특성 등으로 독특한 재료를 이용한 향토 음식으로 발전되어 왔다. (윤숙자 2000)

음청류는 주류 이외에 취미나 기호성 음료를 전반적으로 칭하는 것으로 한국에서의 음청류는 조리 재료와 이것을 만드는 법에 따라 차와 탕, 화채 및 식혜, 미수, 수정과 그리고 제호탕 등으로 구분되며, 술류는 널리 알려진 대로 과일과 곡물의 익힌 것 등을 발효시켜 알콜 성분이 있게 만든 음료의 총칭이다.

농림 축산 식품부 산하 공공 기관인 한식 진흥원 (www.hansik.org)에서는 한식을 주식

류, 찬품류, 양념, 고명 등 4가지로 분류를 하고 있다.

한식 진흥원의 한식 분류 방식에 따르면, 주식류는 밥이나 죽, 만두 및 떡국, 국수 그리고 미음 및 응이 등으로 분류를 하였고, 찬품류는 국이나 탕, 찌개나 전골, 볶음, 찜과 구이, 선, 생, 채 혹은 나물, 조림, 초, 전유어, 적, 회, 쌈과 편육, 족편과 튀각, 부각과 포, 장아찌 및 김치 그리고 젓갈 등으로 분류를 하였다.

양념으로는 소금 및 설탕, 꿀과 조청 그리고 옻, 겨자, 식초, 고추 및 후추, 기름 및 깨소금 등으로 분류를 하였고, 고명은 은행, 알고명, 호두, 알쌈, 버섯, 봉오리(완자), 잣, 황화채, 미나리 초대, 고추, 미나리 등으로 분류를 하고 있다.

앞선 설명에서 보듯이 한식 분류의 복잡함과 학자와 기관에 따라 통일 되지 않은 방식으로 한식 분류가 진행되는 점이 최근에 들어서 부각 되었으며, 한식의 분류를 통일적으로 만들려고 하는 시도가 있었다. (정연경 외 2011)

한식 분류의 복잡성을 탈피하고 통일성을 가지기 위해 한식을 도서 문헌 체계를 따른 분류를 원용하여 사용하기 시작하였다. 이러한 방법으로 한식 분야의 한국 십진 분류법 (Korean Decimal Classification, KDC)과 듀이 십진 분류법 (Dewey Decimal Classification, DDC) 분류 체계가 나오기도 했다.

이상과 같이 한식의 유형은 학자나 관련 기관마다 구분하는 방식과 사용하는 용어가 서로 상이하며, 통일된 규칙이 없이 주로 주식과 부식으로 크게 구분하였고, 주식에는 밥, 죽, 미음, 국수, 만두, 떡국 등이 공통적으로 해당되고, 부식류는 국, 찌개, 전골, 조림, 김치, 찜, 구이, 전, 나물, 장아찌, 회, 편육, 족편, 젓갈 등으로 구분하고 있으며, 자료에 따라 떡 및 한과, 음청류, 주류 등은 부가적인 항목으로 제시하고 있다 (정연경 외 2011)

이처럼 한식은 학자에 따라 또는 사용하는 기관에 따라 한식이 분류되는 기준이 다르고 그 분류 방식에 있어서도 통일된 방식이 없다고 할 수 있다.

또한 이러한 복잡한 한식의 분류 방식이 실질적으로 한식의 세계화를 진행하는 있어서 하나의 어려움으로 진행되었을 것으로 사료되어지며, 또한 이렇게 복잡한 한식 분류와 통일되지 않은 한식의 분류 체계는 본 연구에서도 식수 예측을 진행하는 데 있어서 많은 제약 요소로 작용이 된 것도 사실이며 많은 어려움이 있었다고 할 수 있다.

과거 조선 시대부터 지금까지의 한식의 분류 관련하여 언급한 것에 대해 주의 깊게 살펴 보면, 한식의 분류 방식을 진행하는 있어서 중요한 공통점을 발견할 수 있다는 점이다.

본인은 한국 음식은 결국 어떤 재료를 가지고 어떤 조리 방식으로 음식을 만들었는지에 따라 그 시대의 분류 체계가 달라지고 있다는 점을 발견할 수 있으며, 한식 분류는 결론적으로 조리 재료와 조리 방식에 따라 기본적인 분류를 진행할 수 있다는 것을 알게 해주는 중요한 시사점을 발견해낼 수 있었다.

본 연구를 통해, 저자는 기본적으로 음식이 그 시대에 따른 문화와 환경에 따라 여러 가지 모습으로 변화되고 발전되어 가지만 근본적으로 조리 재료와 조리 방법을 기본으로 하여 한식 메뉴를 분류를 하게 되면 통일성과 확장성을 가질 수 있다는 장점이 있을 뿐 만 아니라 체계적으로 한식 분류법을 만들어서 식수 예측 모델링에 적용할 수 있다는 것을 알 수 있었다.

많은 기관이나 학자들이 한식을 분류할 때 기본적으로 조리 재료와 조리 방법을 혼용하여 진행을 하는 경우 이는 통일성을 가지지 못하게 되는 요인이 되며 또한 확장성에 큰 제약 요소로 작용하게 될 뿐 아니라 복잡성을 가지게 된다.

또한 최근에는 한식 분야의 DDC (Dewey Decimal Classification) 분류 수정 전개 방안이 나왔으며, 이에 따르면, 주식, 분식 및 밀반찬, 후식 및 간식, 음주류로 크게 구분하였으며 한식의 유형과 종류를 세분화하고 있다. (정연경 외 2011)

한식 분야의 DDC (Dewey Decimal Classification) 분류 수정 전개 방안에 따르면, 주식

아래에는 비빔밥을 포함한 밥, 죽, 미음과 국수 만두 떡국으로 구분을 하였으며 국수 만두에는 만두국과 수제비를 분류하였으며, 부식에는 마른 반찬과 포를 분류하고 하위에 국·탕·찌개·전골, 볶음·찜·선·초·조림, 구이·전·적·튀김, 순대·편육·족편, 나물·선채·숙채, 회, 장류·젓갈·장아·자반, 김치 등 9개 항목으로 세분화하였다.

간식 및 후식은 호두과자나 황남빵 등 우리나라의 빵이나 전통 과자를 분류하는 항목으로 하위에 음청류, 떡류 그리고 과정류·과절류 등으로 구분하였고, 떡류에는 약식이 과정류에는 한과, 약과, 산자, 강정을 음청류에는 식례와 수정과 화채를 분류하였고, 주류는 탁주와 증류주로 구분하였으며, 탁주에는 막걸리와 청주를 증류주는 소주를 분류하였다. (정연경 2011)

통일 되지 않고 복잡한 한식 분류 기법에서 탈피하여 새로운 시각으로 한식 메뉴의 분류를 적용하고자 한다. 본 연구에서는, 한식 분야의 DDC (Dewey Decimal Classification) 분류 수정 전개 방안에 따라 크게 한식을 분류하고, 조리 재료와 조리 방법을 기본으로 하여 한식 메뉴를 분류하고자 한다.

본 연구에서 적용한 한식 분류 방법은 기존의 연구와의 큰 차별점으로서 데이터 과학을 기반으로 하는 머신 러닝을 기법을 사용한 점도 있지만 복잡한 한식 메뉴를 단순화하여 식수 예측의 모델링에 적용하는 것이 중요한 차별점이라고 할 수 있다. (정연경 외 2011 ; 김세연 2007 ; 이원갑 외 2014 ; 윤숙자 2000)

향후 3장에서 자세히 설명되는 본 연구에서 적용한 한식 메뉴 분류 방법을 식수 예측 모델링을 적용할 수 있도록 한식의 복잡한 분류를 단순화하고 통일화 시킨 점에 대해서는 아직 까지 국내에서는 시도된 바가 없으며, 최초로 진행이 된 것으로 판단되어지며, 이러한 한식의 분류 기준은 앞으로 한식 분야에 있어서 많은 시사점이 있을 것으로 사료된다.

제 7 절 본 연구의 차별성

본 연구의 차별성은 먼저 메뉴 다양성 및 메뉴 선호도를 고려한 수요 예측 모델링을 하고, 이를 기계 학습 기법을 적용하여 식수 예측을 수행하는 새로운 접근을 시도했다는 것이다.

국내에서 진행된 식수 예측 관련 연구 사례를 살펴보면, 대부분 회귀 분석 등 전통적인 통계 기법을 활용하고 있다.

그러나 식수 인원과 이에 영향을 주는 요인들 간의 인과성을 규명하여 식수 인원 예측에 적용하는 인과형 예측법의 하나인 다중 회귀 분석은 다양한 설명 변수들로 인해 변수들 사이에 강한 상관 관계가 발생하여 다중 공선성이 존재하거나 모델링을 진행하는데 모델 그 자체가 지나치게 복잡해지는 문제가 발생한다는 단점이 있다.

그리고 식수 인원 예측 품질의 중요한 요인인 메뉴 다양성 및 메뉴 선호도를 고려한 수요 예측 모델링을 하고, 이를 기계 학습 기법을 적용하여 식수 예측을 한 사례는 아직 없는 것으로 파악된다. (김혜영 2000 ; 정라나 2001 ; 임혜진 2008 ; 임재영 2015)

둘째, 본 연구에서는 집단 급식소의 실데이터를 사용한 식수 예측을 진행했다는 점이다. 집단 급식소의 식수 예측은 지금까지 여러 차례 설명을 한 바와 같이 매우 중요하다고 할 수 있으며, 이를 위해서는 여러 가지 제약이 있는 전통적인 통계 기법 외 기계 학습 등 다각적인 방법을 이용한 식수 예측에 관한 연구가 보다 활발히 이루어져야 하며, 실데이터를 통해 진행하는 것이 중요하다고 할 수 있다.

셋째, 다양한 예측 모델링을 시도했다. 예를 들어 다중 회귀 모델링 기법 뿐만 아니라 회귀 계수 축소를 통해 보다 모델을 Robust하게 만들어 주는 Ridge 회귀나 영향력이 거의 없는 회귀 계수 값을 0으로 만들어 줌으로서 0이 아닌 변수들 선택에 있어서의 해석력과

설명력을 확보한 분석 방법인 Lasso를 사용하여 식수 예측 모델링을 진행한다.

또한 여러 개의 Decision Tree를 만든 후 최빈값을 기준으로 예측하거나 분류하는 알고리즘인 Random Forest 기법, Bagging과 Boosting 기법 등의 기계 학습 기법을 사용하여 식수 예측 모델링을 진행한다.

그리고 식수 인원 예측을 위해 식수 예측 모델링을 진행하여 실제의 식수 인원 수와 본 연구에서 개발한 식수 예측 인원수와의 오차율을 비교 분석한다.

넷째, 식수 예측의 정확도를 높이기 위해 진행된 식수 예측 모델링의 다양한 기법의 예측 결과를 비교 적용하여 이를 최종적으로 결합하여 새로운 성능이 좋은 식수 예측 모델링을 개발하는 2단계 접근법을 시도하였다.

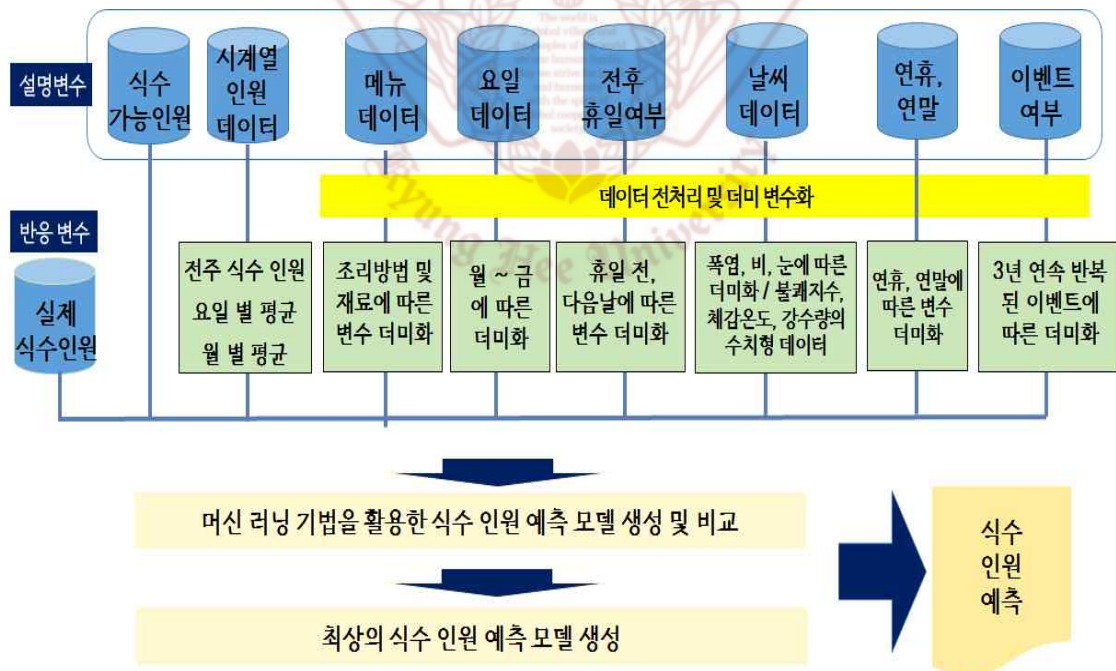
요약하면 본 연구와 같이 실제 메뉴의 선호도 등을 반영하여 기계 학습 기법을 활용하여 식수 예측을 진행한 것은 국내에서는 지금 까지 전례가 없는 것으로 판단되며 (김혜영 2000 ; 정라나 2001 ; 임혜진 2008 ; 임재영 2015), 또한 집단 급식소에서의 식수 예측 관련하여 실무적으로 사용될 수 있도록 기계 학습 기법을 사용하여 식수 인원 예측 정확도를 높임으로써 집단 급식소의 경험에 의한 식수 인원 예측에서 데이터 과학에 기반한 식수 인원 예측으로의 터닝 포인트를 만드는데 기여를 하였다고 본다.

제 3 장 연구 방법

제 1 절 연구 모형

본 연구에서는 실제적으로 집단 급식소의 식수 예측 모델링을 하기 위해 S 시청에서 제공한 데이터와 날씨 등 8개 카테고리에서 생성된 63개의 설명 변수들을 데이터 전처리하여 식수 예측 모델링을 연구 진행을 하였으며, 본 연구의 예측 모델링의 연구 모형은 그림 3-1과 같다

그림 3-1 식수 예측 모델 개발 연구 모형



식수 예측을 위해서 식수 가능 인원 변수, 시계열 인원 변수, 메뉴 특성 변수, 요일 변수,

날씨 변수, 이벤트 변수, 연휴 연말 변수, 전후 휴일 여부 변수 등 8개 카테고리의 63개의 변수를 사용하여 예측 모형을 수립하였다. 단, 미세 먼지가 심한 날의 경우 미세 먼지가 식수 예측에 미치는 영향도 고려하고자 하였으나, 미세 먼지 관련 데이터를 제공하는 기관으로부터 미리 확보 가능한 미세먼지 예보일이 1일치 밖에 없어 단지 하루 빨리 예측을 하는 것은 큰 의미가 없어 실제 식수 인원 예측 모델링에서는 미세 먼지를 설명 변수로 사용하는 것은 제외를 시켰다.

63개의 변수들의 관련 데이터를 가지고 데이터 전처리를 하였으며, 전처리된 데이터를 적용하여 기계 학습 기법을 활용하여 식수 예측 모델을 생성하였고, 생성된 1차 모델링의 식수 인원 예측치와 실제 식수 인원 수와의 오차율에 대해 각각을 비교를 하였다.

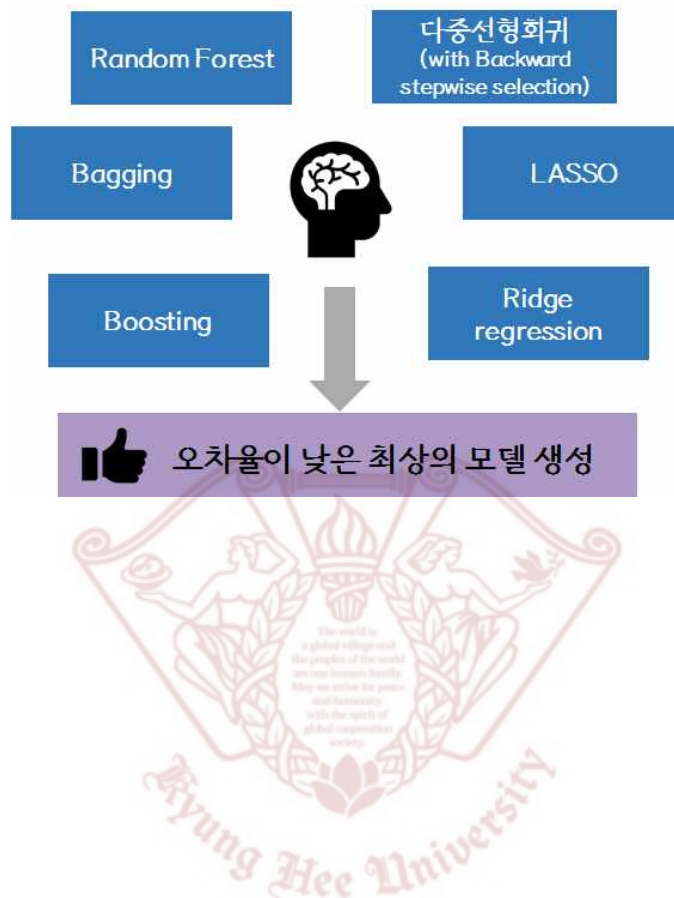
그리고 1차 모델링 시에 Training Set과 Test Set을 구분하여, 3년치의 데이터 중 70%에 해당 하는 데이터를 Training Set으로, 그리고 나머지 30%에 해당하는 데이터를 Test Set으로 진행을 하였다.

또한 무작위로 Training Set과 Test Set 추출을 반복한 Sample에 의한 Bias를 없애는 노력을 기울였다.

식수 예측 모델링 방법으로 총 6개의 모델링 기법을 생성하여 비교 분석을 하였다. R을 활용한 선형 회귀 모델은 3가지로, 다중 선형 모델 (Backwards Stepwise Selection), 능형 회귀 (Ridge Regression) 그리고 LASSO 기법을 적용하였으며, 기계 학습 모델은 Random Forest, Bagging, Boosting 등 3가지 기법을 적용하여 진행 하였으며, 이 6개의 모델을 비교 평가하여 상위 4개의 모델의 결과치의 평균을 내어 최종 모델링을 진행 하여 가장 성능이 좋은 예측 모델을 최종 생성 하였다.

본 연구에서는 회귀 모델과 기계 학습을 활용한 6가지의 모델링 기법을 토대로 식수 예측 알고리즘을 제작하였는데 이를 그림으로 나타내면 그림 3-2과 같다.

그림 3-2 식수 예측 모델링 관련 적용 기법 및 최적의 예측 모델 생성



제 2 절 연구 대상

본 연구는 S시 시청 구내 식당에서 제공한 메뉴, 식사 가능 인원 그리고 날씨 등의 데이터 3년치 (2015년 5월 1일 ~ 2018년 4월 30일 기준)를 대상으로 연구를 진행하였다.

앞선 문헌 연구를 통해 식수 예측에 영향을 미치는 요인과 변수들을 1차적으로 파악하였으며, 그리고 보다 중요한 변수를 발견해 내기 위해 S시청 구내 식당에서 제공한 메뉴, 식사 가능 인원 등의 실제 데이터와 날씨 등의 실제 데이터 3년치를 대상으로 하여 중요 변수 파악을 진행하였으며, 이를 바탕으로 식수 예측 모델링을 진행 하였다.

또한 본 연구에서는 듀이 십진법 분류법에 의한 수정 개정안을 바탕으로 메뉴를 크게 주식, 부식 및 밑반찬, 후식 및 간식 등으로 나누고, 메뉴에 있는 개별 음식을 조리 재료와 조리 방법에 2가지로 분류하여, 1차적으로 조리 방법에 의한 분류를 진행하였으며, 2차적으로는 재료에 의한 분류를 하여 진행을 하였다.

한식 메뉴 분류는 조리 재료와 조리 방식에 따라 듀이 십진법 분류법에 의한 수정 개정안을 큰 틀로 한 음식 메뉴 분류를 바탕으로 S시청에서 제공한 메뉴를 가지고 실제적으로 본 연구에 적용하였다.

본 연구는 741개의 관측치 (2015년 5월 1일 ~ 2018년 4월 30일)를 사용하여 요일, 날씨, 기온, 메뉴 등 8가지의 변수 중에서 R 프로그래밍을 활용하여 식수 예측 모델링을 하였으며, 분석 기법은 다중회귀 분석(Backward Stepwise Regression), 능형 회귀 (Ridge Regression), LASSO, Random Forest, Bagging, Boosting 등 회귀 분석 모델링과 기계 학습 기법을 적용하여 각각 개발된 모델에 대한 예측력을 비교 평가하고 비교 평가된 모델을 기반으로 향후 미래의 데이터 예측의 안정성을 위해 예측력이 높은 최적의 모델을 생성 진행하였다.

제 3 절 활용 데이터

본 연구의 이용 가능한 데이터로 서울시 구내 식당 3년치 사용 내역 데이터와 기상청의 날씨 데이터를 식수 인원 예측 모델링을 위해 사용하였다.

음식물 잔반을 줄이기 위해 본 연구의 대상으로는 S 시청 구내 식당에서 매일 버리는 음식량을 의미하는 일일 잔반량 (일평균 435L, 연간 110,900L, 2015년 10월 12일 ~ 2016년 10월 11일 기준)으로 정해서, 이를 대상으로 음식물 잔반을 줄이기 위해 식수 예측 모델링을 진행하였다.

적정 식수 모델링을 위해 사용한 공공 데이터는 다음과 같다.

가. S시청 구내 식당 이용 데이터 3년치

- 일자별 식사 인원
- 일자별 식단 메뉴
- 일자별 식사 이용 가능자 현황 (일자별 현원 및 출장, 휴가자 수 등)

나. 날씨 관련 데이터 3년치

- 일자별 날씨 및 온도 현황

본 연구 대상으로 활용한 실데이터는 2015년 5월 1일부터 2018년 4월 30일까지의 공공 데이터를 가지고 진행하였으며, 그 내용은 표 3-1과 같다.

표 3-1 식수 예측 모델링을 위해 사용한 공공 데이터

구분	보유 기관	설명	요약	기준 시점
직원 식당 메뉴 및 실제 이용 인원	S시	S시 식단 메뉴 및 실제 식수	식단 메뉴 및 실제 식수	2015년 5월 1일~ 2018년 4월 30일 (3년치)
식수 가능 인원	S시	S시 본청 일자별 현원 및 출장, 휴가자수	날짜, 본청 현원, 행정국 예외 직원, 출장자, 휴가자, 식수 가능자	
날씨 데이터	기상청	강수 형태, 강수량, 기온, 풍속, 습도	날짜 및 시간, 지역, 강수 형태, 기온, 풍속, 습도	

한편 식수 예측 모델링 관련하여 사용한 변수들에 대한 내용을 표로 요약하면 표 3-2와 같다.

표 3-2 실제 식수 모델링 사용 변수 내용 요약

설명 변수	변수 내용	반응 변수
식수 가능 인원변수	휴가자, 출장자 수 고려	실제 식수
시계열 인원 변수	전주 식수인원, 요일 평균 인원, 월 평균 인원	
요일 및 계절 변수	월 화 수 목 금 및 봄 여름 가을 겨울	
메뉴 특성 변수	식단의 조리 재료 및 조리 방법 고려	
전날 다음날 휴일 변수	전날 다음날이 휴일 여부	
날씨 변수	불쾌지수 체감온도 폭염 눈 비 강수량	
이벤트 변수	3년 연속 반복된 이벤트의 유무	
연휴 연말 변수	연속 3일 이상 휴일 전날 혹은 다음날, 연말 (12월21일 ~ 31일)	

반응 변수인 실제 식수 인원을 위한 식수 예측 모델링을 위해 S시와 기상청에서 제공 받아 만든 변수 그리고 Feature Engineering에 의한 변수 생성 등을 통해 적용한 데이터를 요약 정리하면 표 3-3과 같다

표 3-3 실제 식수 모델링을 위한 가용 데이터 요약

구분	보유 기관	설명	제공 컬럼명 요약	기준 시점
본관 직원식당 이용인원	S시	서울시 집단 급식소 현황	날짜, 날씨, 기온, 식단, 인원, 잔반	2015년 5월1일 ~ 2018년 4월30일
본청 일자별 현원 및 출장, 휴가자 수	S시	본청 식사 가능자	날짜, 본청 현원, 행정 국 예외 직원, 출장자, 휴가자, 식사 가능자	
날씨 데이터	기상청	기온, 하늘의 흐림 정도, 비/눈/진눈깨비 유무	기온, 하늘상태, 불쾌 지수, 체감온도, 폭염, 눈, 비, 강수량	
시계열 인원 변수	자체 생성	일정한 주기와 특성을 가진 인원 변수	전주 식수 인원, 요일 평균 식수 인원, 월 평균 식수	
연휴 연말 변수	자체 생성	전날 다음날 휴일 여부	전후 휴일 여부	
이벤트 변수	자체 생성	3년 동안 반복된 이벤트 여부	복날, 무료 떡국, 대보름 등	
연휴 연말 변수	자체 생성	연휴의 전날 및 다음날 여부와 연말 표기	연속된 3일 이상 휴일, 연말(12월21일~31일)	
메뉴 특성 데이터	S시	3년치 점심 메뉴 현황	식단 내용	

제 4 절 구성 변수

식수 예측에 미치는 변수 파악을 위해 크게 내부적 요인과 외부적 요인 변수 등 2가지 요인으로 구분을 하였다.

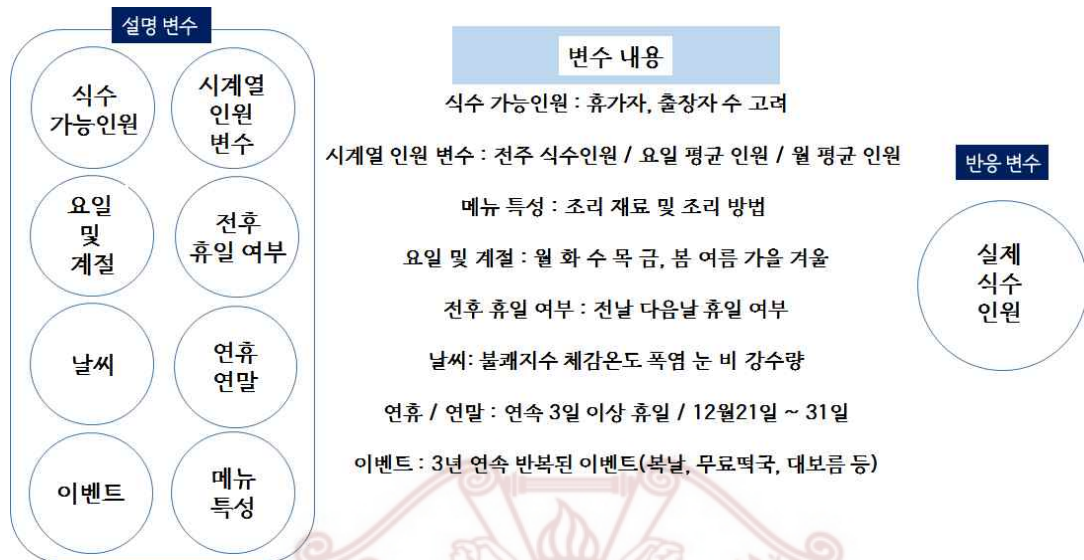
내부적 요인으로 고려를 할 변수로는 일일 메뉴, 당일 이용 가능 인원 (출근 인원, 출장자, 휴가자 등), 전주 식사 인원, 요일 평균 식사 인원, 월 평균 식사 인원, 일일 식수 가능 인원 등을 고려하여 분류를 하였다.

외부적 요인으로는 날씨, 기온, 전날 혹은 다음날에 휴무일/공휴일 여부, 요일, 이벤트 여부 연휴 및 연말 여부 등을 고려하여 분류를 하였다.

식사 인원 예측에 영향을 미치는 변수 파악을 위해 내부적 요인과 외부적 요인 등을 고려한 변수들을 8개의 카테고리로 정리하면 아래와 같으며, 이를 그림으로 나타내면 그림 3-3과 같다.

- ① 식사 가능 인원
- ② 시계열 인원 변수
- ③ 메뉴 특성 변수
- ④ 연휴 연말 변수
- ⑤ 이벤트 여부 변수
- ⑥ 요일 및 계절 변수
- ⑦ 전날과 다음날 휴일 여부 변수
- ⑧ 날씨 및 기온 변수

그림 3-3 식수 예측 모델링에 사용한 변수 요약



식수 예측 모델링을 위해 개개인의 식사 여부를 결정함에 있어 영향을 미칠 수 있는 가능한 여러 요인들을 모아 모델링을 진행을 하였다.

특히 공공 기관의 집단 급식소라는 특성을 감안하여 식당을 이용할 수 있는 인원의 대상자인 식수 가능 인원, 그리고 메뉴, 날씨 및 기후, 환경 변수, 시계열 인원 변수 등의 8개 카테고리의 63개 변수를 사용하여 식수 예측 모델링을 진행하였다.

제 5 절 구성 변수의 상세 내용

1. 식사 가능 인원 (당일 식사 이용 가능 인원)

가. 식수 가능 인원 선정

S시의 구내 식당의 당일 이용 가능 인원을 예측하기 위해, S시로부터 받은 과거 3년치의 출근 인원, 출장자, 휴가자 등을 고려한 식사 가능자의 데이터를 사용하였으며, 구하는 식은 아래와 같다.

식수 가능 인원 수 = 본청 현원 - 행정국 예외 직원 - 본청 출장자 - 본청 휴가자

위에서 언급한 기준으로 식수 가능 인원을 파악하여 일별 숫자를 조사하였으며, 이에 대한 자료는 표3-4와 같다.

표 3-4 식수 가능 인원 일일 현황 파악표

대상일	요일	본청 현원 (A)	국 예외 직원 (B)	본청 출장자 (C)	본청 휴가자 (D)	본청 식사가능자 (A-B-C-D)
20151012	월	2263	249	179	143	1692
20151013	화	2263	249	228	144	1642
20151014	수	2263	249	216	158	1640
20151015	목	2263	249	228	182	1604
20151016	금	2263	249	355	184	1475
20151019	월	2263	249	252	123	1639
20151020	화	2263	249	259	129	1626
20151021	수	2263	249	224	131	1659

현원 추출 기준으로 인사 DB는 월별로 관리되어, 월별로 현원 추출하였으며, 일자별 전입/전출 사항은 반영되지 않았다. 공무원, 준공무원 등이 300여명이 소속되어 있는 국이

있는데 이 조직의 1/2 정도는 본청 외에 다른 청사에도 근무하고 교대 근무자가 대부분
이므로 이 중의 1/2이 식사 가능자로 구분하고 대략적으로 본청 식사 가능자는 이중 1/4
이라고 판단하였다.

출장자는 식사 불가능자라고 판단하였으며, 출장자의 기준은 다음과 같은 기준으로 하였
다. 출장일이 2일 이상인 경우 해당 기간은 식사 불가능이라고 판단하였으며, 출장이 1일
이면서 출장 시작 시각, 종료 시각 둘 다 식사 기간 (12~13시) 전 또는 후인 경우를 제
외한 모드는 건을 산정하였으며 출장 시작/종료 시각이 12시 ??분 인 경우는 식사 불가
능자로 판단하였다.

휴가자 또한 식사 불가능자라고 판단하였으며, 연가, 병가, 특별 휴가, 대체 휴무, 장기
휴가 등 복무 관련 시스템 상 모든 휴가를 포함하였으며, 오전/오후 반가는 모두 식사를
할 수 없다고 판단하였다.

또한 외부에서 온 사람의 식사 인원을 파악을 할 수가 없어 외부에서 와서 식사를 한
인원은 식사 가능 인원내 편입하여 진행을 하였으며, S시의 본청 근무 부서 기준은 일정
시점의 본청 근무 부서 목록을 기준으로 하였으며, 대상 기간 중 발생한 조직 개편은 반
영되지 않았다.

또한 미래의 식수 인원을 예측하기 위해 사용되는 식수 가능 인원은 현재 시점에서 파악
할 수 없으므로 과거의 3년치 식수 가능 인원 데이터를 기본으로 예측을 하여 예측하고
자 날의 본청 식사 가능자를 월별과 요일별로 평균을 내어서 미래의 식수 가능 인원으로
활용하였다.

식수 예측 모델링 개발 후 7일간의 식사 인원을 미리 예측하기 위해서 미래의 식사 가능
인원을 파악 할 수가 없어서, 과거 3년치 자료를 바탕으로 본청 식사 가능자의 월별 요일
별 평균을 구해 식사 가능 인원의 예측하여 식수 가능 인원으로 사용하였다.

3년 동안 S 시청의 구내 식당에서는 2015년 5월 1일 ~ 2018년 4월 30일 기간 동안 월

요일은 141주, 화요일은 156주, 수요일은 147주 목요일은 150주, 금요일은 147주 동안 식사가 제공이 되었고, 식사 가능 인원은 대상 기간 동안 일별 식사 가능자를 요일 별로 분류하여 요일별 식사 가능자의 평균을 구해보면 표 3-5와 같다.

표 3-5 요일별 식수 가능 인원 현황표

월	화	수	목	금
1735	1709	1685	1676	1614

위의 표 3-5에서 보듯이 식수 가능 인원은 주중 1610 ~ 1730 여 명 수준으로 파악되었으며, 요일 중 월요일에 식수 가능 인원이 제일 많았으며 금요일이 제일 식사 인원이 적은 것을 알 수 있다.

또한 식수 가능 인원이 제일 많은 월요일과 제일 적은 금요일의 차이는 약 120여명으로 파악되었으며 요일별 평균 식수 가능 인원을 그림으로 나타내면 그림 3-4와 같다

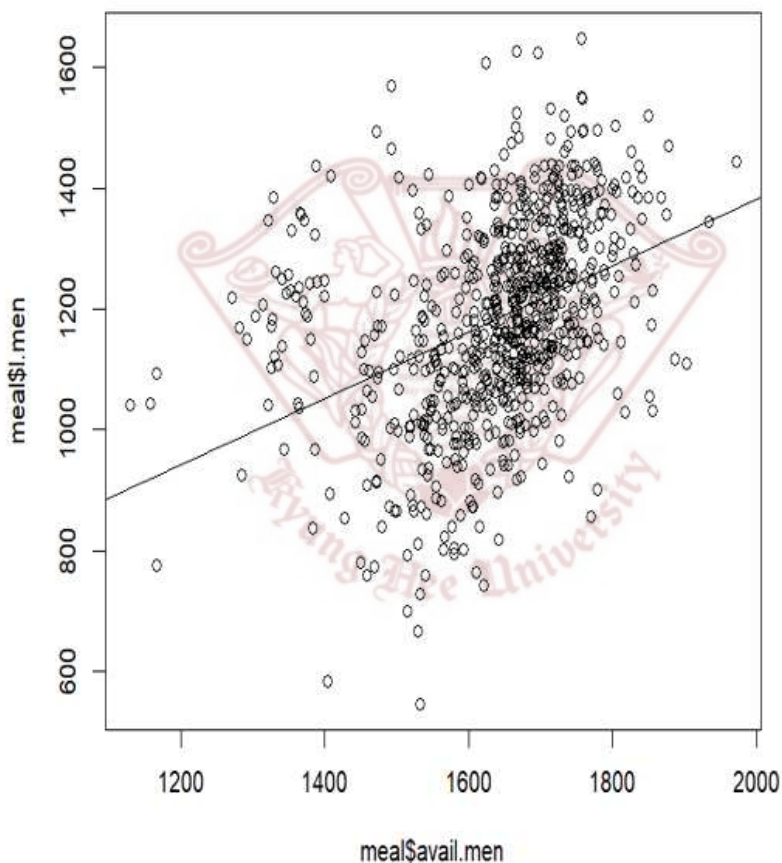
그림 3-4 요일별 평균 식수 가능 인원



나. 식수 가능 인원과 실제 식수 인원의 관계 분석

식수 가능 인원을 설명 변수로 하여 실제 식수 인원과 식수 가능 인원의 관계를 파악하기 위해 단순 회귀를 적용하였으며, 실제 식수 인원과 식수 가능 인원의 관계를 나타내는 산포도는 그림 3-5와 같다

그림 3-5 실제 식수 인원과 식수 가능 인원의 상관 관계



또한 실제 식수 인원과 식수 가능 인원의 상관 관계 파악을 위해 단순 회귀를 적용한 결과 식수 가능 인원 변수의 Adjusted R²의 값이 0.1491로 식수 가능 인원 변수의 설명력이 약 14.9%이었다 (p-value : $< 2.2e-16$).

Call:

```
lm(formula = l.men ~ avail.men, data = meal)
```

Residuals:

Min	1Q	Median	3Q	Max
-578.1	-100.8	1.6	107.4	465.6

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	286.56829	78.65323	3.643	0.000288 ***
avail.men	0.54700	0.04785	11.431	< 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 155.2 on 739 degrees of freedom

Multiple R-squared: 0.1502, Adjusted R-squared: 0.1491

F-statistic: 130.7 on 1 and 739 DF, p-value: < 2.2e-16

식수 가능 인원 변수는 실제 식수 예측에 실질적인 영향을 미치는 변수로 파악이 되었으며, 미래의 식수 가능 인원의 데이터는 미리 알 수가 없으므로, 미래의 식수 인원을 예측하는 경우에는 식수 가능 인원의 예상치로 과거 식수 가능 인원의 3년치를 기본으로 월 별 요일 별 평균을 내어 진행을 하였다.

2. 시계열 인원 변수

식수 예측에 영향을 미치는 변수를 파악하기 위해 현장 조사를 수행하였다. S시의 집단 급식소의 경우에 영양사 및 조리사는 경험에 의해 식수 인원을 파악하여 진행을 하고 있었는데, 요일 별 식수 인원을 주요 결정 요인으로 인식하고 있었다. 식수 인원은 요일 별로 일정한 패턴이 있는 것으로 파악하고 있었다.

또한 집단 급식소에서 식수 예측의 오차율에 영향을 미치는 분석이 있었는데, 집단 급식소에서의 식수 예측시 고려하는 요인으로 요일, 제공 메뉴 선호도, 날씨, 전주 동일 요일 식수, 하루 운영 끼니, 고객 행사, 1달 이전 식수 인원 자료, 계절, 동일 메뉴 과거 식수, 이벤트, 전년 해당월 혹은 해당 주의 식수, 기온, 주변 식당 이용 가능성, 동일 메뉴 과거 잔반량 등으로 파악하고 있다. (임재영, 2015)

이러한 상황 속에서 식수 예측에 있어서 과거의 경험치 데이터는 중요하다는 것을 알 수 있으며, 이를 활용하는 것을 식수 예측에 중요한 변수가 된다는 것을 알 수 있다.

본 연구에서도 이에 착안하여, 과거의 식수 데이터를 확보하여 식수 예측 모델링에 적용을 하였으며, 시기별 혹은 주기별로 일정한 패턴을 보이고 있는 인원 변수를 시계열 인원 변수라 명칭하고 분석 및 모델링을 진행을 하였다.

본 연구에서 사용한 시계열 인원 변수는 전주 식사 인원, 요일 별 식수 인원과 월별 식수 변수를 Feature Engineering 기법으로 새로 생성하였으며, 요일별 식수 인원은 요일 평균으로 그리고 월별 식수 인원은 월 평균을 구하여 각각 요일 평균과 월 평균 변수로 사용하여 식수 예측 모델링을 진행을 하였다

Feature Engineering 기법에 의한 새로운 변수 생성 관련하여 요약한 결과는 다음과 같다.

① 요일 별 식수 인원

현장에서 경험치로 식수 인원을 예측함에 있어 중요한 요인으로 삼고 있는 것은 요일별 식수 인원이었다. 이는 현장에서 요일별로 다르게 식수 인원을 파악하고 식사 준비를 진행하고 있다는 점에 착안하여 본 연구에서도 요일별 식수 인원을 변수를 생성하여 진행을 하였다

② 전주 식수 인원

본 연구에서는 또한 요일별 식수 인원 변수 외에 추가적으로 식수 인원 예측 모델링이 정확도에 기여할 변수를 찾아 요일별 식수 인원 변수에 추가적으로 보완을 해줄 변수로써, 요일이 일치하는 같은 요일의 전주 식사 인원을 파악하여 전주 식수 인원이라는 변수를 추가적으로 생성하여 진행을 하였으며, 이로 인해 식수 인원 예측 모델링의 정확도가 향상되었다.

이를 통해 요일별 식수 인원과 요일별의 전주 식수 인원을 동시에 고려함으로써 모델링의 오차율을 줄이고 정확도를 더욱 높일 수 있는 계기를 마련하였다.

③ 월 별 식수 인원

또한 1년 중 월별로 식수 인원이 항상 고르게 분포하는 것이 아니라 연말과 같이 월별로 식수 인원이 달라 같은 요일이라고 하더라도 월에 따라 다르다는 것을 파악하여, 월 별 식수 인원이라는 변수를 생성하여 요일별 식수 인원, 전주 식수 인원 과 함께 월 별 식수 인원 변수를 추가적으로 진행을 하였으며, 이 3가지 변수를 추가함으로 인해 식수 인원 예측 모델링의 설명력이 더욱 올라가게 되었다.

Feature Engineering을 하기 전 (요일별 식수인원, 전주 식수 인원, 월 별 식수 인원 변수 제외)의 모델링의 설명력은 다음과 같다.

Call:

```
lm(formula = l.men ~ ., data = meal)
```

Residuals:

Min	1Q	Median	3Q	Max
-538.94	-61.76	1.11	64.28	288.83

Residual standard error: 102.9 on 684 degrees of freedom

Multiple R-squared: 0.6539, Adjusted R-squared: 0.6255

F-statistic: 23.07 on 56 and 684 DF, p-value: $< 2.2e-16$

이 시계열 변수 3가지를 제외한 경우와 3가지 변수를 포함하여 전체 진행한 다중 회귀 모델링 식의 설명력의 차이는 약 4%로써 3가지 변수를 제외한 식의 모델링의 설명력은 0.6255이었으며, Feature Engineering에 의해 생성된 3가지 변수 등을 포함한 전체 모델링의 설명력은 0.6647이었다.

가. 전주 식사 인원

1) 전주 식사 인원 변수 생성

S사에서 제공한 전주의 실제 식수 수를 사용하여 표 3-6과 같이 같은 요일의 전주 식수 인원이라는 새로운 변수로 만들어서 이 변수가 얼마나 설명력이 있는지를 단순 회귀 모형을 통해 분석을 진행하였다.

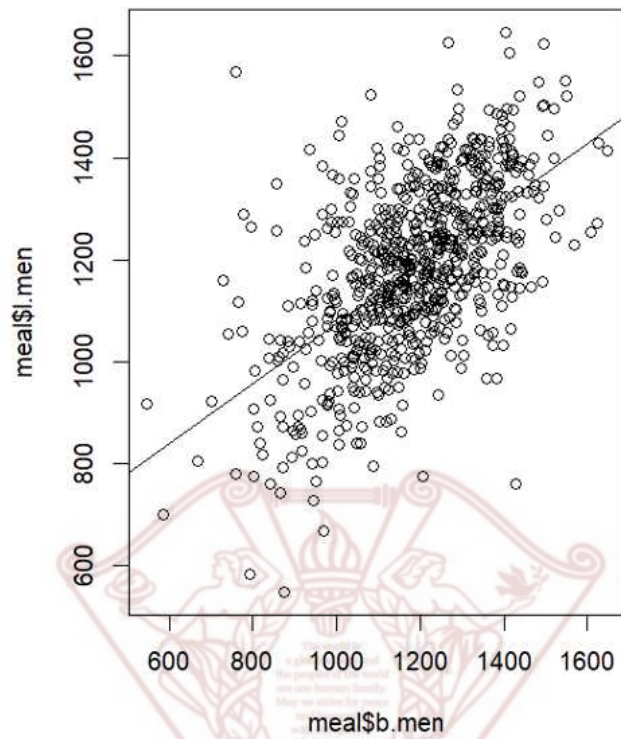
표 3-6 전주 식수 인원 생성표

대상일	요일	실제 식수	전주 식수
20180402	월	1274	1385
20180403	화	1120	1136
20180404	수	1300	1100
20180405	목	1160	1070
20180406	금	818	824
20180409	월	1333	1274
20180410	화	1130	1120
20180411	수	1133	1300

실제 식수 인원과 식수 가능 인원의 상관 관계를 나타내는 산포도는 그림 3-6과 같다

그림 3-6을 실제 식수 인원과 식수 가능 인원의 상관 관계를 나타내는 그림 3-5와 비교하여 보면, 전주 식사 인원 변수가 식수 가능 인원 변수 보다 값들이 회귀선에 더 집중화되어 있는 것을 알 수 있다.

그림 3-6 실제 식수 인원과 전주 식수 인원의 상관 관계



2) 실제 식수 인원과 전주 식수 인원의 상관 관계 분석

3년치 데이터를 가지고 전주 식수 인원을 설명 변수로 한 단순 회귀식의 Adjusted R2의 값이 0.3417로 전주 식사 인원 변수의 설명력이 약 34.2%이었으며 또한 이 변수의 P Value가 낮아 (p-value: $< 2.2e-16$) 전주 식사 인원 변수는 의미가 있는 변수로 파악되었다.

Call:

```
lm(formula = l.men ~ b.men, data = meal)
```

Residuals:

Min	1Q	Median	3Q	Max
-567.64	-90.01	-0.98	87.19	634.88

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	487.73795	35.79695	13.62	$< 2e-16$ ***
b.men	0.58734	0.02993	19.62	$< 2e-16$ ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 136.5 on 739 degrees of freedom

Multiple R-squared: 0.3426, Adjusted R-squared: 0.3417

F-statistic: 385.1 on 1 and 739 DF, p-value: $< 2.2e-16$

또한 식수 가능 인원 변수의 기울기는 0.547이나, 전주 식수 변수의 0.587로 그림 3-6을 실제 식수 인원과 식수 가능 인원의 상관 관계를 나타내는 그림3-5과 비교하여 보면, 전주 식사 인원 변수의 회귀선이 식수 가능 인원 변수의 회귀선 보다 우상향하고 있음을 확인할 수 있다. 그러므로 식수 가능 인원의 변수 보다 전주 식수 변수의 설명력 이 더 크다는 것을 확인 할 수 있다.

나. 요일별 식수 (요일 평균 변수)

1) 요일별 식수 변수 생성

S사에서 제공한 실제 식수 수를 사용하여 요일 별 식수 인원의 평균을 구해서 요일 평균이라는 변수로 새로 만들어 이 변수가 얼마나 설명력이 있는지를 단순 회귀 모형을 통해 분석을 진행하여 파악하였다.

3년 치의 실제 식수 인원을 조사한 결과, 월요일 식수 인원이 제일 많고 금요일은 식수 인원이 제일 적었으며 화 수 목의 식수 인원은 비슷한 식수 인원수를 보이는 등 요일 별로 일정한 패턴이 있었다.

실제 현장에서 확인을 해보니 현장에서도 요일 별로 식수 인원이 다르다는 것을 파악하여 경험치로 월 화 수 목 금 등 요일 별로 식수 인원을 다르게 예측하여 활용하고 있었는데, 이는 실제 식수 파악에 있어서 요일이 중요한 변수라는 것을 의미한다고 할 수 있다.

S사의 실제 식수 인원의 요일별 평균치는 표 3-7과 같다

표 3-7 요일 별 식수 (요일 평균) 인원

요일	월	화	수	목	금
평균 식수	1336명	1208명	1212명	1151명	1001명

평균적으로 월요일이 1,336명으로 실제 식수인원이 제일 많았으며, 금요일은 1,001명으로 제일 작았다. 통상적으로 월요일은 주말을 쉬고 나서 근무하기 때문에 외출 등이 적은 것으로 판단되어지며, 또한 휴일 다음 날에는 될 수 있는 대로 사무실에서 근무를 하는 분위기가 형성되어 있는 것을 반영하고 있는 것으로 판단된다.

그리고 금요일이 월요일에 비해 약 330명 가량 적은 것은 금요일이 채식의 날로 제공 되는 음식이 채식 위주의 음식이어서 채식 음식에 대한 선호도가 떨어지고 있다는 것을

반영한다고 할 수 있으며, 또한 다음 날이 휴일이어서 외출이나 식사를 외부에서 하는 분위기가 형성되어 있는 것을 반영하고 있는 것으로 판단된다. 이를 그래프로 나타내면 그림 3-7과 같다.

그림 3-7 요일별 평균 식수 인원



S시에서 제공한 데이터를 기본으로 하여 실제 식수 인원 수를 사용하여 요일별 식수 인원을 평균값을 내어 요일 평균의 변수를 표 3-8과 같이 생성하여 식수 인원 예측 모델링을 진행하였다.

표 3-8 요일 별 식수 (요일 평균) 인원 생성표

대상일	요일	요일 평균
20180420	월	1336
20180421	화	1208
20180422	수	1212
20180423	목	1151
20180424	금	1001
----	----	----

2) 실제 식수 인원과 요일별 평균 식수 인원의 상관 관계 분석

요일 별 평균 식수 인원 (요일 평균 변수)을 설명 변수로 한 단순회귀식의 Adjusted R의 값이 0.4181로 요일 평균 변수의 설명력이 약 41.8%이었으며 또한 이 변수의 P Value가 낮아 (p-value: $< 2.2e-16$), 의미가 있는 변수로 파악되었으며 그 결과는 다음과 같다.

Call:

```
lm(formula = l.men ~ 요일 평균, data = meal)
```

Residuals:

Min	1Q	Median	3Q	Max
-579.41	-76.50	7.25	83.26	415.26

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-7.65452	51.81486	-0.148	0.883
요일평균	1.00770	0.04366	23.080	$< 2e-16$ ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 128.3 on 739 degrees of freedom

Multiple R-squared: 0.4189, Adjusted R-squared: 0.4181

F-statistic: 532.7 on 1 and 739 DF, p-value: $< 2.2e-16$

3년 치 데이터를 가지고 회귀 분석을 하였을 때, 식수 가능 인원 변수의 Adjusted R²의 값이 0.1491로 식수 가능 인원 변수의 설명력이 약 14.9%이었으며 전주 식수 인원을 설명 변수로 한 Adjusted R²의 값이 0.3417로 전주 식사 인원 변수 설명력이 약 34.2%이었던 값에 비해, 요일평균 변수의 Adjusted R²의 값이 0.4181로 요일 평균 변수의 설명력은 41.8%로 가장 설명력이 높은 변수로 확인 되었다.

다. 월 별 식수 인원 변수 (월 평균 변수)

1) 월 별 식수 인원 변수 생성

S시에서 제공한 실제 식수 수를 사용하여 월 별 식수 인원의 평균을 구해서 월 평균이라는 변수로 새로 만들어 이 변수가 얼마나 설명력이 있는지를 단순 회귀 모형을 통해 분석을 진행하여 파악하였다.

3년치의 데이터로 월별 실제 식수 인원을 조사하였더니, 연말인 12월에 실제 식수 인원의 수가 제일 적었으며, 2월에 식수 인원이 제일 많았으며, 월별로 최대치와 최소치 사이에 약 225명의 차이를 보이고 있었다.

S시의 실제 식수 인원의 월별 평균치는 표 3-9와 같다.

표 3-9 월 별 식수 (월 평균) 인원

월	1월	2월	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월
평균	1206	1276	1209	1160	1145	1161	1221	1201	1181	1204	1188	1051
식수	명	명	명	명	명	명	명	명	명	명	명	명

S시에서 제공한 실제 식수 인원 수를 사용한 데이터를 기본으로 하여 월별 식수 인원 평균을 내어 표 3-10과 같이 월 평균의 변수를 생성하여 분석을 진행하였다.

표 3-10 월 별 식수 (월 평균) 인원 생성표

대상일	요일	월 평균
20180420	월	1160명
20180421	화	1160명
20180422	수	1160명
20180423	목	1160명
20180424	금	1160명
----	---	----

2) 실제 식수 인원과 월 별 식수 인원의 상관 관계 분석

월 별 평균 식수 인원을 설명 변수로 한 단순 회귀식의 Adjusted R의 값이 0.09305로 단순 회귀식의 설명력이 약 9.3%이었으며 또한 이 변수의 P Value가 낮아 (p-value: $< 2.2e-16$), 의미가 있는 변수로 파악되었으며 그 결과는 다음과 같다.

Call:

```
lm(formula = l.men ~ 월평균, data = meal)
```

Residuals:

Min	1Q	Median	3Q	Max
-634.4	-101.5	5.5	113.5	445.5

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-2.4636	135.3284	-0.018	0.985
월평	1.0025	0.1143	8.770	$<2e-16$ ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 160.2 on 739 degrees of freedom

Multiple R-squared: 0.09427, Adjusted R-squared: 0.09305

F-statistic: 76.92 on 1 and 739 DF, p-value: $< 2.2e-16$

3년치 데이터를 가지고 회귀 분석을 하였을 때, 식수 가능 인원 변수의 Adjusted R²의 값이 0.1491로 식수 가능 인원 변수의 설명력이 약 14.9%이었으며 전주 식수 인원을 설명 변수로 한 Adjusted R²의 값이 0.3417로 전주 식사 인원 변수 설명력이 약 34.2%이었으며, 요일평균 변수의 Adjusted R²의 값이 0.4181로 요일 평균 변수의 설명력은

41.8%이었던 것에 비해, 월평균 변수의 Adjusted R2의 값이 0.09305로 월평균 변수의 설명력이 약 9.3%으로 가장 낮은 설명력을 보이는 것으로 확인되었다.

앞선 설명과 같이 식수 가능 인원 및 시계열 인원 변수 등 인원 관련한 변수 중 설명력이 가장 높은 변수는 요일별 식수 인원인 요일 평균 변수, 그 다음이 전주 식수 인원 변수, 그 다음이 식수 가능 인원 변수 그리고 마지막으로 월별 식수 인원인 월 평균 변수임을 확인 할 수 있었으며, 이를 표로 나타내면 표 3-11과 같다.

표 3-11 단순 회귀 방식을 통한 인원 관련 변수의 설명력 요약

변수	Adjusted R2	변수의 설명력	비교
요일 평균 변수	0.4181	41.8%	요일별 평균 식수 인원
전주 인원 식수 변수	0.3417	34.2%	전주의 실제 식수 인원
식수 가능 인원 변수	0.1491	14.9%	식사 가능 이용자
월 평균 변수	0.09305	9.3%	월별 평균 식수 인원

3. 메뉴 특성 변수

가. 한식 메뉴 분류 체계 수립

한국의 음식 문화에 대한 정보 및 분류 체계는 문헌 분류 체계와 달리 분류 체계가 복잡하고 통일되어 있지 못하고 있어서 정보의 활용과 접근에 있어 실제 사용상에서 여러 가지 제약이 있는 상황이라고 할 수 있다.

특히 과거에서부터 한식 관련하여 한식의 정의 및 유형, 분류 체계의 미비, 연구 자료의 절대 부족, 한식 이미지의 부재, 한식 메뉴 외국어 표기법의 비표준화 등이 지속적으로 여러 문제점으로 지적되기도 하였다. (이민아, 2008)

그래서 최근에 들어와서 이러한 한식 메뉴의 분류 체계 관련 미비점을 보완하고자 많은 노력을 하여 왔으며, 이러한 노력의 일환으로 한식 음식 메뉴의 분류를 위해 한국 십진 분류법 (Korean Decimal Classification, KDC) 과 듀이 십진 분류법 (Dewey Decimal Classification, DDC) 같은 문헌 분류 체계를 한식의 분류 체계에 접목을 하는 시도를 하여 왔다.

문헌 분류 체계로 사용되고 있는 한국 십진 분류법 (KDC)은 비정기적인 개정 주기로, 용어의 최신성과 실용성이 떨어지는 단점이 있어 사용상에서 점점 멀어지고 있으나, 듀이 십진 분류법 (DDC)가 주기적이며 지속적인 용어 확장을 통해 최신성과 실용성을 유지하기 때문에 듀이 십진 분류법 (DDC)은 문헌 분류 체계로 도서관에서 뿐 만 아니라 인터넷 자료를 분류하는 기반으로 사용하고 있다. (정연경 외 2010)

또한 문헌 분류 체계로서 세계적으로 많이 사용되고 있는 듀이 십진 분류법 (DDC)은 서구 중심의 분류표로서 음식 분류 체계로 사용되는 경우에는 동양 음식의 유형과 특징을 제대로 반영하지 못한다는 단점이 있다. (정연경 외 2011)

그러므로 국내외적으로 한식에 대한 정보를 조직하고 효율적으로 검색하기 위해서는 이러한 한식 메뉴 분류 체계에 전 세계적으로 가장 많이 활용되고 있는 문헌 분류 체계의 장점을 가지고 와서 한식에 관한 내용을 반영하는 것이 중요하다.

이를 위해 서구 중심의 분류표로서 동양 음식의 유형과 특징이 제대로 반영되지 못한다는 단점을 극복하면서 전 세계적으로 가장 많이 활용되는 듀이 십진법 (DDC)의 수정 개정안을 사용하여 한식 메뉴의 분류 특성을 담는 방안이 제시되기도 하였다. (정연경 외 2011).

그러나 한국 음식 메뉴에 대한 분류에 대해서는 학계와 관계 기관에서 공통적으로 사용할 수 있게 하기 위해 계속 논의 중이며, 아직 까지 체계화되고 통일화된 방법이 없으며, 학자와 관련 기관에 따라 그 분류 방법이 다르다고 할 수 있다.

또한 이렇게 복잡한 한식 메뉴 분류와 통일되지 않은 한식의 분류 체계는 본 연구에서도 식수 예측 모델링을 위한 메뉴 특성을 반영한 프로그래밍을 진행하는 데 있어서 많은 제약 요소로 작용이 된 것도 사실이며 실질적으로 많은 어려움이 있었다고 할 수 있다.

이에 본 연구에서는 한국 음식이 기본적으로 결국 어떤 재료를 가지고 어떤 조리 방식으로 음식을 만들었는지에 따라 그 시대의 분류 체계가 달라지고 있다는 점에 주목하고자 한다.

또한 한식을 분류하는 많은 기관이나 학자들이 기본적으로 조리 재료와 조리 방법을 혼용하여 진행을 하는 경우 이는 통일성을 가지지 못하게 되며 또한 확장성에 큰 제약 요소로 작용하게 될 뿐 아니라 복잡성을 가지게 된다는 점을 파악하였다. (정연경 외 2011 ; 김세연 2007 ; 이원갑 외 2014 ; 윤숙자 2000) 이는 한식 분류는 조리 재료와 조리 방식에 따라 기본적인 분류를 진행할 수 있다는 중요한 시사점을 가진다.

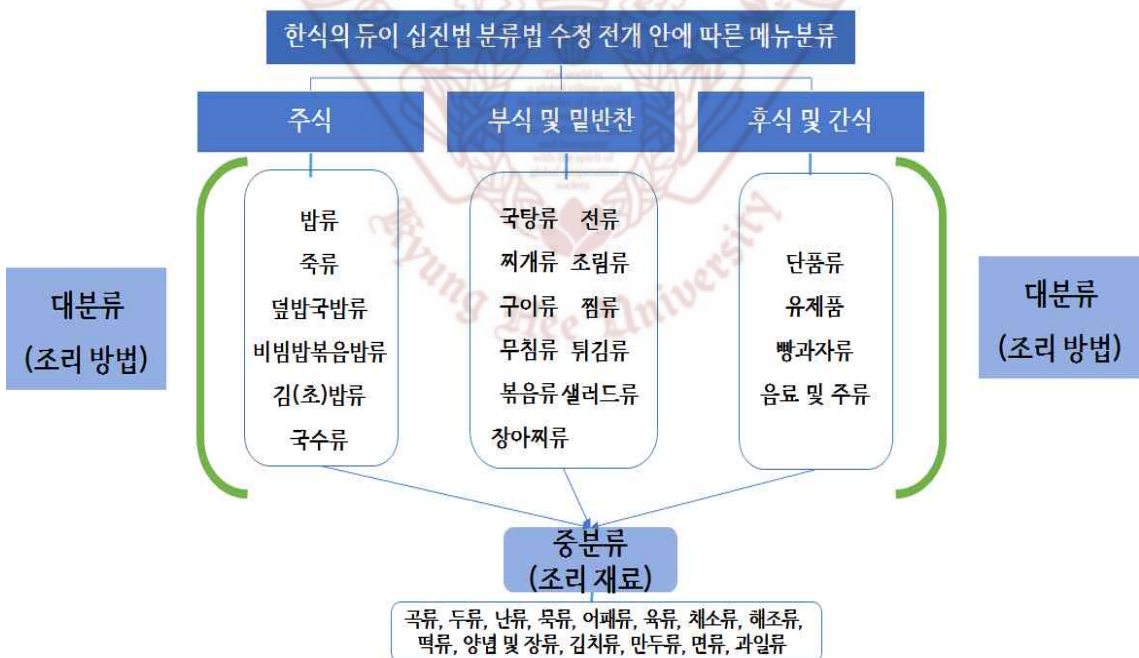
따라서 본 연구에서는 듀이 십진법 분류법에 의한 수정 개정안을 큰 틀로 하여 한식 음식

식 메뉴 분류에 적용을 하였다. 두이 십진법 분류법에 의한 수정 개정안을 바탕으로 메뉴를 크게 주식, 부식 및 밑반찬, 후식 및 간식 등으로 나누고, 메뉴에 있는 개별 음식을 조리 재료와 조리 방법에 2가지로 분류하여, 1차적으로 조리 방법에 의한 분류를 진행하였으며, 2차적으로는 재료에 의한 분류를 하여 진행하였다.

그러므로 본 연구는 어떤 특정 음식에 대해 어떤 재료들이 합쳐져서 어떤 조리 방법으로 만들어졌는지를 기본으로 진행하였으며, 두이 십진법 분류법에 의한 수정 전개안을 큰 틀로 삼아 재료의 형태와 조리 방식에 따라 구분하는 독창적인 메뉴 분류 체계를 만들었다.

이를 그림으로 나타내면 그림 3-8과 같다

그림 3-8 한식 메뉴의 두이 십진법 분류법 수정 개정안을 기본으로 한 메뉴의 분류 체계



나. 한식 메뉴의 분류 기준

한식 메뉴 관련하여 실데이터를 처리하는 데 있어서 사용되고 있는 일일 식단의 메뉴를 처리하고 적용하는 프로그래밍을 개발하는 것은 쉽지 않은 사항이었으며, 한식의 특성상 복잡하고 다양한 메뉴를 분류하고 코딩하기 위해서는 한국 음식에 대한 분류의 체계를 세우는 것이 매우 중요하다.

집단 급식소의 식수 예측 알고리즘을 만드는데 있어서 실제 메뉴를 적용할 수 있도록 데이터를 전처리하고 이를 활용하여 진행할 수 있도록 한 메뉴의 분류 기준 수립은 식수 예측 알고리즘 개발을 위해 중요한 부분이다.

본 연구에서는 한식 음식 메뉴 분류 체계에 있어서 우선적으로 공신력 있는 국가 기관의 분류 기준을 참고하고, 향후 확장성을 고려하여 최종적으로 집단 급식소에 맞게 조리 방법 및 재료를 잘 분류할 수 있는 기준인 한식 분야의 듀이 십진법 분류법 (Dewey Decimal Classification)의 수정 개정안을 참고로 하여 분류를 하였으며, 한식 분야의 듀이 십진법의 수정 개정안의 내용을 요약하면 표 3-12과 같다.

표 3-12 한식 메뉴의 듀이 십진법 분류법의 수정 개정안에 따른 분류 체계 요약

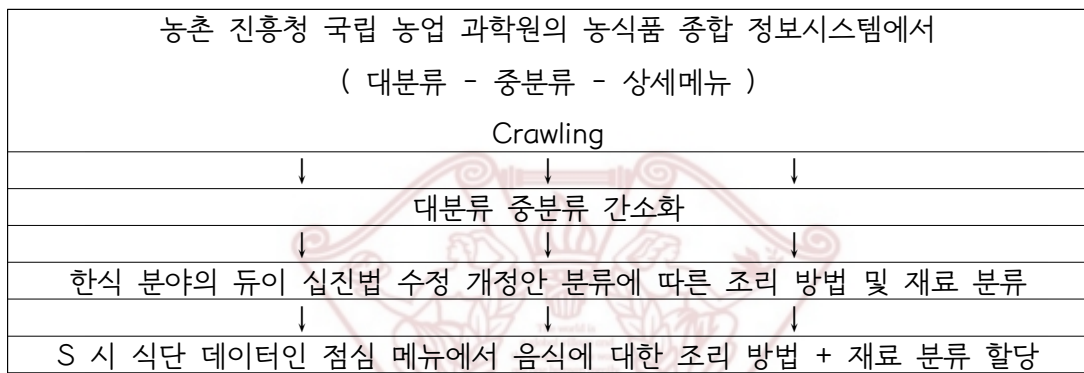
주식	부식 및 밑반찬	후식 및 간식	음주류
밥, 죽, 국수 만두, 수제비 떡국	국이나 탕, 찌개나 전골, 찜 혹은 조림, 구이나 볶음, 튀김, 순대, 나물, 전, 장류, 젓갈, 김치 등	떡류, 과점류, 음청류	탁주, 증류주

또한 이러한 듀이 십진법 분류법의 수정 전개 방안에 따라 기본적으로 메뉴를 크게 주식, 부식 및 밑반찬, 후식 및 간식 등으로 나누고, 메뉴에 있는 개별 음식을 메뉴 재료와 조리 방법에 2가지로 분류하여, 메뉴에 있는 특정 음식에 대해 어떤 재료들이 합쳐져서 어떤 조리 방법으로 만들어졌는지를 데이터화 하여 이를 데이터 전처리 통해 모델링을 진행

하였다.

본 연구는 국가 기관인 농림 수산부 산하 농촌진흥청의 국립 농업 과학원의 한국 음식 메뉴 분류 기준을 기반으로 하여 한식 메뉴의 분류상의 중복 여부를 제거하고 이를 바탕으로 유효하게 코딩할 수 있도록 한국 음식 메뉴의 분류를 진행하였으며, 이를 표로 정리 하면 표 3-13과 같다.

표 3-13 본 연구에서 적용한 한식 음식 메뉴의 분류 절차 요약



본 연구에서 참고한 한국 음식 메뉴 분류 기준이 된 사이트는 농림 수산부의 국가 기관인 농촌 진흥청의 국립 농업 과학원에서 운영하는 공신력 있는 사이트인 농식품 종합 정보 시스템 사이트로서, 이 사이트에서는 전통 향토 음식과 전통 향토 음식의 문화, 식품 영양, 기능성 정보, 식단 관리 등의 정보들을 주로 제공하고 있다.

본 연구에서는 우선적으로 농촌 진흥청 산하 국립 농업 과학원의 농식품 종합 정보 시스템의 식단 관리 메뉴젠 (<http://koreanfood.rda.go.kr/kfi/mgnmenumkFoodSelect/list>) 사이트의 대분류, 중분류, 음식명을 참고 하였다.

국립 농업 과학원의 농식품 종합 정보 시스템의 식단 관리 메뉴젠은 그림 3-9와 같이 크게 대분류, 중분류 그리고 음식명으로 분류하고 있다.

메뉴젠의 대분류는 총 23가지로 분류되어 있으며, 그 종류는 밥류, 빵 과자류, 면 만두류, 죽류, 국탕류, 찌개류, 찜류, 구이류, 전류, 볶음류, 조림류, 튀김류, 무침류, 김치 짬아찌, 회류, 젓갈류, 양념류, 유제품, 음료 및 주류, 과일류, 단품류, 떡류, 기타 등 23가지로 구성되어 있다.

그림 3-9 국립 농업 과학원의 농식품 종합 정보 시스템 식단 관리 메뉴젠

음식분류

대분류

중분류

음식명

Q 검색

• 검색 결과 총 1186개의 음식정보가 검색되었습니다.

선택	대분류	중분류	음식내용	총량
<input type="checkbox"/>	밥류	쌀밥	누룽지	50g
<input type="checkbox"/>	밥류	쌀밥	누른밥	50g
<input type="checkbox"/>	밥류	쌀밥	쌀밥	90g
<input type="checkbox"/>	밥류	쌀밥	찰밥	90g
<input type="checkbox"/>	밥류	쌀밥	현미밥	90g

또한 메뉴젠의 중분류는 54가지로 분류되어 있어, 그 종류는 쌀밥, 잡곡밥류, 채소밥류, 비빔밥볶음밥류, 덮밥국밥류, 김(초)밥, 빵류, 면류, 만두류, 죽류, 맑은국류, 된장국류, 곰국탕류, 냉국류, 기타, 어패류 찌개, 육류 찌개, 된장 찌개류, 전골류, 어패류찜, 육류찜, 채소류찜, 어패류 구이, 육류구이, 어패류전, 육류전, 채소류전, 어패류 볶음, 육류 난류 볶음, 채소류 볶음, 곡류 두류 볶음, 어패류 조림, 육류 난류 조림, 채소류 조림, 두류 기타 조림, 어패류 튀김, 육류 튀김, 채소 해조류 튀김, 숙채, 어패류 무침, 육류 무침, 김치류, 어패류회, 육류회, 채소류회, 젓갈류, 장아찌, 양념류, 종합 양념, 유제품, 음료 및 주류, 과일류 및 그 제품, 단일 식품 자체, 떡류 등 54가지로 구성되어 있다.

메뉴젠은 대분류 23가지와 중분류 54가지 등 총 77 가지의 카탈리고리로 분류가 되어 있었다.

국립 농업 과학원의 농식품 종합 정보 시스템의 식단 관리 메뉴젠을 자세히 보면, 메뉴젠의 메뉴 분류인 대분류와 중분류의 내용 중 겹치는 것이 있어, 그대로 사용하는 하는 경우 문제가 있어 이를 듀이 십진법 분류법에 의한 수정 전개안에 따라 겹치는 것을 조정하여 요리 재료와 요리 방법에 따라 재분류를 해서 사용하여야 한다.

그러므로 농촌 진흥청 산하 국립 농업 과학원의 농식품 종합 정보 시스템인 식단 관리 메뉴젠의 분류에 따르면 조리 재료와 재료 방법이 대분류와 중분류에도 혼재되어 나타나고 있어, 본 연구에서는 이에 따른 비효율성을 제거하기 위해 대분류는 조리 방법을 기준으로, 중분류는 조리 재료를 기준으로 분류를 하여 재편성하였다.

또한 실제 현장에서 추가 되는 메뉴의 향후 확장성 등을 고려하여 메뉴 가지 수가 대분류 23가지, 중분류 54가지 등 총 77가지를 듀이 십진법 분류법에 의한 수정 전개안에 따라 요리 재료와 요리 방법에 따라 재분류하여 대분류는 조리 방법에 따라 21가지 그리고 중분류는 조리 재료에 따라 18가지 등 총 39가지로 수정 보완하였다.

메뉴젠의 대분류 23가지 중 면만두류를 중분류의 면류, 만두류로 대체하였으며, 김치 짬아찌를 중분류의 김치류, 짬아찌류로 변경하고, 회류를 중분류의 어패류로 변경하였으며, 젓갈류는 S시의 집단 급식소 메뉴에서 없어서 삭제하였고, 양념류를 양념 및 장류로 이름 변경하고, 기타는 추가 확장시 고려하는 것으로 하여 여기서 삭제하였다.

메뉴젠의 대분류는 조리 방법을 기준으로 조정을 하였으며, 18가지로 최종 정리된 대분류는 표 3-14와 같다

표 3-14 농식품 종합 정보 시스템의 식단 관리 메뉴젠의 간소화된 대분류

간소화된 대분류			
밥류	단품류	떡류	구이류
국탕류	빵과자류	찜류	죽류
전류	찌개류	조림류	튀김류
양념 및 장류	유제품		

농식품 종합 정보 시스템의 식단 관리 메뉴젠의 중분류 54가지 중 쌀밥 잡곡밥류 채소밥류를 쌀밥으로 통합하고, 빵류는 대분류의 빵과자류로 통합하였고, 맑은국류, 된장국류, 곰국탕류, 냉국류를 대분류의 국탕류로 통합하였고, 어패류 찌개, 어패류 찜, 어패류 구이, 어패류 전, 어패류 볶음, 어패류 조림, 어패류 튀김, 어패류 무침, 어패류 회 등 9가지는 어패류로 통합하였다.

그리고 된장찌개류, 전골류를 대분류의 찌개류로 통합하였고, 육류찌개, 육류찜, 육류구이, 육류전, 육류 튀김, 육류무침, 육류회 등 7가지는 육류로 통합하였으며, 곡류 두류 볶음과 두류 기타 조림을 곡류/두류로 분리하고, 채소류전, 채소류 볶음, 채소류 조림, 채소류회 등 4가지는 채소류로 통합하였다.

또한 육류난류 볶음과 육류난류 조림은 육류/난류로 각각 분리하고, 채소 해조류 튀김은 채소류/해조류로 분리하였으며, 양념류와 종합 양념류를 대분류의 양념 및 장류로 통합하고, 과일류 및 그 제품을 대분류의 과일류로 통합하였으며, 단일 식품 자체를 대분류의 단품류로 통합하였고, 기타와 젓갈류는 제외하였다.

본 연구에서는 최종적으로 농촌진흥청 산하 국립 농업 과학원의 농식품 종합 정보 시스템인 식단 관리 메뉴젠의 중분류 54가지는 조리 재료를 기준으로 21가지로 조정을 하였으며, 21가지로 최종 정리된 중분류는 표 3-15와 같다.

표 3-15 농식품 종합 정보 시스템의 식단 관리 메뉴젠의 간소화된 중분류

간소화된 중분류			
쌀밥	만두류	음료 및 주류	어패류
비빔밥볶음밥류	곡류	김치류	채소류
덮밥국밥류	두류	장아찌	해조류
김(초)밥	떡류	육류	목류
죽류	유제품	난류	샐러드류
면류			

본 연구에서는 한식 분야의 듀이 십진법 분류법 수정 개정안의 큰 틀에서 농촌 진흥청 산하 국립 농업 과학원의 농식품 종합 정보 시스템인 식단 관리를 참고하여 조리 재료와 조리 방법에 따른 집단 급식소의 한식 메뉴의 분류 체계를 적용하였으며, 이를 요약 정리 하면 표 3-16과 같다.

표 3-16 한식의 듀이 십진법 분류법 수정 개정안에 따른 메뉴의 분류

한식의 듀이 십진법 분류법에 따른 메뉴 분류			
주식	부식 및 밑반찬	후식 및 간식	대분류 (조리 방법에 따른 분류)
밥류	국탕류, 찌개류		
죽류	구이류, 무침류	단품류	
덮밥국밥류	볶음류, 장아찌류	유제품	
비빔밥볶음밥류	전류, 조림류	빵과자류	
김(초)밥류	찜류, 튀김류	음료 및 주류	
국수류	샐러드류		
↘	↓	↙	
중분류 (조리 재료)			중분류 (조리 재료에 따른 분류)
곡류, 두류, 난류, 묵류, 어패류, 육류, 채소류, 해조류, 떡류, 양념 및 장류, 김치류, 만두류, 면류, 과일류			

S 시청의 실제 메뉴를 바탕으로 식수 예측 알고리즘을 개발하는 데 있어서, 실제로 현장에서 사용하고 있는 메뉴의 분류 체계를 데이터 과학에 기반한 기계 학습 기법의 컴퓨터 프로그래밍에 적용할 수 있도록 한 것이 지금 까지 다른 논문에서 시도되지 않은 본 연구에서의 중요한 차별점 중 하나라고 할 수 있다.

다. S시 집단 급식소 메뉴의 적용

본 연구는 앞서 설명한대로, 한식 분류의 듀이 십진법 수정 개정안을 큰 틀로 하여 농촌진흥청 산하 국립 농업 과학원의 농식품 종합 정보 시스템 메뉴젠의 식단 관리를 참고하여, 조리 방법과 조리 재료에 의해 분류 체계를 진행하였다.

음식 메뉴의 특징을 선별하여, 조리 방법과 재료 등 크게 두가지 특성에 의해 분류하였으며, 대분류로 조리 방법에 따라 삶거나 구이류, 국탕류, 김치류, 단품류, 무침류, 밥류, 볶음류, 빵과자류, 양념류, 음료 및 주류, 장아찌류, 전류, 조림류, 죽류, 찌개류, 찜류, 튀김류 등 국립 농업 과학원의 대분류를 21가지 방법으로 수정 재분류하였다.

중분류로 재료에 따라 곡류두류, 곰국탕류, 과일류, 기타, 김(초)밥, 난류, 냉국류, 덮밥국밥류, 된장국류, 떡류, 만두류, 맑은국류, 면류, 묵류, 부침류, 비빔밥볶음밥류, 샐러드류, 숙채, 쌀밥, 양념류, 어패류, 유제품, 육류, 잡곡류, 채소류, 해조류 등 국립 농업 과학원의 중분류를 18가지로 수정 재분류하여 S시로부터 제공 받은 식단의 실제 데이터를 적용하여 진행을 하였다.

표 3-17은 S시로 부터 제공 받은 식단 데이터의 일부이다.

표 3-17 S시로 부터 제공 받은 식단

월	일	요일	날씨	기온	중식단	인원	잔반	비고
10	12	월	맑음	8~18	어묵무국 제육볶음 야채잡채 치커리유자청무침	1419	400	
10	13	화	맑음	8~21	흰떡미역국 갈치감자조림 미니새송이무침 청경채생채	1335	500	
10	14	수	맑음	11~24	자장밥 계란파국 군만두 양장피	1152	400	

					짜사이채무침			
10	15	목	맑음	12~26	맑은콩나물국 춘천식닭갈비 오이고추쌈장무침 양상추샐러드	1130	500	
10	16	금	맑음	13~24	취나물밥&양념장 청국장찌개 고추잡채김말이튀김 표고버섯초무침 얼갈이나물	1095	500	

또한 3년 치의 실제 메뉴는 약 1200여개의 단품 메뉴로 분류 되었으며, 이러한 단품 메뉴의 특성 분류를 한 것을 예시를 하면 표 3-18과 같다.

표 3-18 단품 메뉴 특성 분류 예시

메뉴	중분류 (재료1)	중분류 (재료2)	중분류 (재료3)	중분류 (재료4)	대분류 (조리방법)
가래떡	떡류				단품류
가자미 구이	어패류				구이류
가자미 양념 구이	어패류	양념 및 장류			구이류
가자미 양념찜	어패류	양념 및 장류			찜류
가자미 조림	어패류				조림류
가자미 찜	어패류				찜류
가자미 튀김	어패류				튀김류
가지 나물	어패류				무침류

단품 메뉴의 특성에 따라 표 3-18과 같이 분류를 한 데이터를 다시 전처리하여 메뉴 특성을 반영하는 변수로 생성하여 식수 예측 모델링에 적용하였다.

라. 실제 식수 인원과 메뉴 특성 변수의 상관 관계 분석

S시에서 제공한 실제 식단 메뉴를 사용하여 조리 방법과 조리 재료를 나타내는 39가지의 메뉴 특성을 반영하는 변수를 만들어 전체 메뉴 변수의 설명력을 알기 위해 회귀 분석을 진행하였다.

표 3-19는 식수 예측 모델링에 사용되는 실제 메뉴 분류에 대한 예를 보여주는 것이다.

표 3-19 실제 식단 메뉴 분류 예시

2016년 9월1일(목)	중분류 (재료1)	중분류 (재료2)	중분류 (재료3)	중분류 (재료4)	대분류 (조리방법)
콩가루시래기국	두류	채소류			국탕류
쇠고기버섯볶음	육류	채소류			볶음류
양장피	면류	채소류	해조류		볶음류
도라지생채	채소류				무침류

또한 메뉴의 중요도를 모델에 직접 반영하고자 주메뉴와 중요한 일부 메뉴만 모델링에 반영한 것이 아니라 모든 식단에 제공하는 모든 메뉴를 모델링에 반영을 하였다.

식수를 종속변수로, 39가지의 메뉴 특성 변수를 설명 변수로 한 회귀식의 Adjusted R의 값이 0.2381로 전체 메뉴 특성 변수의 회귀식 설명력이 약 23.8%이었으며 전체 메뉴 특성 변수의 P Value가 $2.2e-16$ 로 낮아 의미가 있는 변수로 파악되었으며 그 결과는 다음과 같다.

Call:

```
lm(formula = l.men ~ ., data = meal)
```

Residuals:

Min	1Q	Median	3Q	Max
-538.03	-88.54	3.28	95.54	462.95

Coefficients: (1 not defined because of singularities)

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1171.440	52.854	22.164	< 2e-16 ***
곡류	24.618	12.483	1.972	0.048993 *
두류	-16.959	12.502	-1.356	0.175383
난류	61.917	18.426	3.360	0.000821 ***
목류	-3.929	20.846	-0.188	0.850552
어패류	37.217	9.719	3.829	0.000140 ***
육류	90.339	10.577	8.541	< 2e-16 ***
채소류	-8.611	6.474	-1.330	0.183903
해조류	-26.111	12.765	-2.046	0.041166 *
떡류	5.188	19.169	0.271	0.786726
양념.및.장류	14.447	9.949	1.452	0.146916
김치류	-3.551	12.011	-0.296	0.767571
만두류	2.910	24.319	0.120	0.904773
면류	20.882	14.045	1.487	0.137516
과일류	20.679	15.675	1.319	0.187508
밥류	-34.112	13.663	-2.497	0.012766 *
죽류	-127.492	59.492	-2.143	0.032453 *
덮밥국밥류	-16.692	31.817	-0.525	0.599996
비빔밥볶음밥류	-39.239	24.578	-1.596	0.110825
김.초.밥류	NA	NA	NA	NA
국수류	-44.338	46.406	-0.955	0.339680
국탕류	-59.390	35.329	-1.681	0.093197 .
찌개류	-52.307	38.045	-1.375	0.169603
구이류	28.240	17.912	1.577	0.115342
무침류	8.294	13.408	0.619	0.536388
볶음류	10.207	15.792	0.646	0.518262
장아찌류	-30.652	28.169	-1.088	0.276893

전류	7.652	19.313	0.396	0.692048
조림류	5.272	17.493	0.301	0.763215
찜류	-8.953	22.118	-0.405	0.685761
튀김류	-25.305	15.754	-1.606	0.108652
샐러드류	28.097	18.230	1.541	0.123690
단품류	-8.609	10.130	-0.850	0.395680
유제품	48.946	30.660	1.596	0.110846
빵과자류	-72.395	38.275	-1.891	0.058975
음료.및.주류	-46.522	32.742	-1.421	0.155800

Signif. codes:

0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 146.8 on 706 degrees of freedom

Multiple R-squared: 0.2731, Adjusted R-squared: 0.2381

F-statistic: 7.802 on 34 and 706 DF, p-value: < 2.2e-16

3년치 데이터를 가지고 회귀 분석을 하였을 때, 식수 가능 인원 변수의 Adjusted R2의 값이 0.1491로 식수 가능 인원 변수의 설명력이 약 14.9%이었으며 전주 식수 인원을 설명 변수로 한 Adjusted R2의 값이 0.3417로 전주 식사 인원 변수의 설명력이 약 34.2%이었으며, 요일평균 변수의 Adjusted R2의 값이 0.4181로 요일 평균 변수의 설명력은 41.8%이었으며, 월평균 변수의 Adjusted R2의 값이 0.09305로 월평균 변수의 설명력이 약 9.3%으로 가장 낮은 설명력을 보이고 있는 것에 비해 39가지의 메뉴 특성의 전체를 반영한 변수의 설명력은 Adjusted R2의 값은 0.2381이었으며, 설명력은 23.8% 이었다.

결론적으로 식수 예측을 위한 주요 변수들 중 설명력이 큰 변수는 요일별 식수 인원인 요일 평균 변수, 전주 식수 변수, 메뉴 특성 변수, 식수 가능 인원 변수, 월별 식수 인원인 월 평균 변수의 순임을 확인 할 수 있었으며, 이를 표로 나타내면 표 3-20과 같다.

표 3-20 주요 변수의 설명력 요약

변수	Adjusted R2	변수의 설명력	비교
요일 평균	0.4181	41.8%	요일별 평균 식수 인원
전주 식수	0.3417	34.2%	전주의 실제 식수 인원
식수 가능 인원	0.1491	14.9%	식사 가능 이용자
월 평균	0.09305	9.3%	월별 평균 식수 인원
메뉴 특성	0.2381	23.8%	39가지의 메뉴 특성 전체



4. 날씨 변수

날씨 데이터는 기상청의 홈페이지의 기상 자료 개방 포털 사이트로 부터 가져 왔다. 기상청으로 부터 받은 자료는 동네 예보 자료로써, 동네 예보는 지역별, 시간별 차이로 인한 수요자의 불편을 최소화하기 위해 전국을 5km×5km 간격의 촘촘한 격자점으로 3시간 마다 읍, 면, 동 단위로 상세한 날씨를 제공하는 예보이다. 격자는 가로(동서) 149개, 남북(세로) 253개로 총 37,697개로, 현재 기상청에서는 도로명 및 행정동 기반으로 예보를 제공하고 있다.

동네 예보 실황은 현재의 기상 실황을 나타내며, 매 시각 일 24회 발표 한다.

- 시행일 : 2010.06.15. 17:00 KST
- 생산 주기/예보 기간 : 매 시각 발표(일 24회)
- 예보 요소 : 1시간 기온, 1시간 강수량, 강수형태, 하늘상태, 풍향, 풍속, 습도, 뇌전(8개)

기상청으로 받은 데이터는 표 3-21과 같다.

표 3-21 기상청 데이터

자료 포맷	<div>기상 자료 개방 포털 : CSV</div> <div><div>- DD (일자), HHmm(시간), Value(값)</div><div>- 예) 기간: 2016년 06월, 요소: 강수형태를 선택한 경우</div></div> <div><div>DATA</div><table><tr><td>1</td><td>0</td><td>0</td></tr></table><div>↓</div><div>해석</div><table><tr><td>1일</td><td>00:00</td><td>강수 없음</td></tr></table></div>	1	0	0	1일	00:00	강수 없음
1	0	0					
1일	00:00	강수 없음					
특정 요소의 코드값 및 범주	<div>하늘 상태 코드 : 맑음(1), 구름조금(2), 구름많음(3), 흐림(4)</div> <div>• 강수 형태 코드 : 없음(0), 비(1), 비/눈(2), 눈(3)</div>						

	※ 비/눈은 진눈깨비를 의미
자료 접근방법	기상청 홈페이지 - http://www.kma.go.kr/weather/main.jsp • 기상 자료 개방 포털→데이터→ 날씨예보→ 실태 분석자료 • 공공 데이터포털→ 오픈API → 동네예보 정보 조회서비스

날씨 (비, 눈, 불쾌지수, 폭염, 체감 온도)가 급식소의 식수 인원에 영향을 미친다고 판단하였고, 기상청의 기상 자료 관련 개방 포털 (<https://data.kma.go.kr>)에서 서울특별시 중구 명동 지역의 2015.5.01 ~ 2018.4.30 기간 동안 데이터 (기온, 강수 형태, 강수량, 풍속, 습도) 를 수집하여 사용하였다.

기온은 섭씨 단위 (° C)로 측정이 되었으며, 강수 형태는 없음 (0), 비 (1), 진눈깨비 (2), 눈(3)으로 표시를 하였으며, 풍속은 (m/s) 단위로 표시되고, 습도는 % 단위로 표시된다

비, 눈, 불쾌 지수, 폭염, 체감 온도 등 날씨 관련 데이터의 전처리는 다음과 같이 진행을 하였다.

비, 눈은 강수 형태 코드가 비(1), 진눈깨비(2)인 경우 rain 변수에 1, 그 외의 경우 0의 값을 부여하고, 강수 형태 코드가 눈(3)인 경우 snow 변수에 1을 부여하고, 그 이외의 경우 0의 값을 부여하였다.

폭염은 기상청에서 폭염의 정의를 33° C 이상인 날로 정의하고 있어, 기상청의 정의에 따라 기온이 33° C 이상이면 1을 부여하고, 그 이외의 경우 0의 값을 부여하였다.

불쾌지수는 기온과 습도의 조합으로 사람이 느끼는 더위를 표현을 하는 것으로 기상청에서 수집한 기온과 습도를 활용하여 불쾌지수를 계산하였다.

체감 온도는 기온과 풍속의 조합으로 사람이 느끼는 추위를 표현을 하는 것으로 기상청

에서 수집한 기온과 풍속을 활용하여 체감온도를 계산하였다.

기상청으로 받은 데이터를 가지고 기온, 하늘 상태, 눈, 비 등에 따라 기온 전처리, 하늘 상태 전처리, 눈 전처리, 비 전처리를 하였으며, 이를 표로 나타내면 표 3-22와 같다.

표 3-22 날씨 데이터 전처리

기온 전처리	<ul style="list-style-type: none"> • 2015년 5월 1일 ~ 2018년 4월 30일 기간 중구의 기온 데이터 수집 • 10~14시의 데이터만 추출한 뒤 max값을 구함 • 봄 (3월 ~ 5월), 여름 (6월 ~ 8월), 가을 (9월 ~ 11월), 겨울 (12월 ~ 2월)에 따라 기준을 둠 • 겨울에는 주의가 필요한 기온인 -5도를 기준으로 삼음 • 여름에는 매우 위험할 수 있는 28도를 기준으로 삼음 • 봄/가을에는 평균 기온이 가장 낮은 값인 3월을 참고하여 5도를 기준으로 삼음
하늘 상태 전처리	<ul style="list-style-type: none"> • 2015년 5월 1일 ~ 2018년 4월 30일 기간 중구의 하늘 상태 데이터 수집 • 10~14시의 데이터만 추출하여 값을 모두 더한 후 평균 냄 • 맑음(1), 구름 조금(2), 구름 많음 (3), 흐림(4)
비 전처리	<ul style="list-style-type: none"> • 2015년 5월 1일 ~ 2018년 4월 30일 기간 중구의 강수 종류 데이터 수집 • 10~14시의 데이터만 추출하여 최빈값을 구함 • 없음(0), 비(1), 진눈깨비(2), 눈(3) • 1은 비, 2는 진눈깨비를 뜻함 • 1과 2일 때 비가 왔다고 판단하여 rain변수에 1을 줌
눈 전처리	<ul style="list-style-type: none"> • 2015년 5월 1일 ~ 2018년 4월 30일 기간 중구의 강수 종류 데이터 수집 • 10~14시의 데이터만 추출하여 최빈값을 구함 • 3은 눈을 뜻함 • 3일 때 눈이 왔다고 판단하여 snow변수에 1을 줌

날씨 데이터 중 비와 눈의 전처리 기준은 표 3-23과 같으며 비와 눈이 온 경우는 각각 더미 변수화 하였다.

표 3-23 비와 눈의 판단 기준

강수 형태	내용 기준
비	비가 오거나 진눈깨비가 오는 경우 비가 온 것으로 간주함
눈	눈이 오는 경우에 눈이 온 것으로 간주함

강수 형태 코드가 비(1), 진눈깨비(2)인 경우 rain 변수에 1을 주고, 그 이외의 경우 0의 값을 주었고 강수 형태 코드가 눈(3)인 경우 snow 변수에 1을 주고, 그 이외의 경우 0의 값을 주었다.

구름의 전처리 관련하여 구름의 정도를 다음의 기준에 따라 더미화 하였다. 하늘 상태 코드가 구름 많음(3)과 흐림(4)일 경우 cloud 변수에 1을 주고 그 이외의 경우 0의 값을 주었다.

또한 이상 기온 (매우 춥거나 더울 때)일 경우, 식수 인원에 영향을 미칠 수 있다고 판단 하였고 기준은 표 3-24와 같다. 날씨 데이터에서 이상 기온에 해당될 경우 temp변수에 1을 주고 그렇지 않을 경우 0의 값을 주었다.

표 3-24 이상 기온의 판단 기준

시기	판단 기준
여름 (6월~8월)	28도 이상 여름철 더위 체감 지수에 따르면 28도 이상인 경우 열사병의 위험이 높음.
겨울 (12월~2월)	-5도 이하 겨울철 체감 온도 지수에 따르면 -5도 이하인 경우 동상의 위험이 있음.
봄(3월~5월), 가을(9월~11월)	5도 이하 추위에 대한 준비가 거의 없어 준비가 해이해질 시기이지만, 추위로 인한 피해를 입을 수 있어 봄철과 가을철의 최저 평균 온도인 5도를 기준으로 삼았음.

이를 기반으로 날씨 관련한 데이터 전처리한 것을 일부 표시하면 표 3-25와 같다. 표3-25에서 2015년 5월 11일 날씨는 비가 오고, 불쾌지수가 68, 체감온도가 23을 의미하고, 2015년 5월 12일 날씨는 불쾌지수가 58, 체감온도가 12임을 의미한다.

표 3-25 날씨 데이터 전처리 결과 일부 예시

일시	불쾌지수	체감온도	폭염	rain	snow
20150511	68	23	0	1	0
20150512	58	12	0	0	0



5. 이벤트 여부 변수

이벤트의 경우에는 비교란에 이벤트가 표시되어 있는 것을 고려하여 이벤트 여부 변수는 이벤트가 있는 날과 이벤트 없는 날로 구분을 하였고, 이벤트 있는 날은 이벤트가 긍정 이벤트인지 부정 이벤트 인지 여부를 조사하였다.

이벤트 당일 식수인원이 전체 식수 평균 보다 많으면 긍정 이벤트로 평균보다 낮으면 부정 이벤트로 분류하였다. 이를 그림으로 나타내면 그림 3-10과 같다.

그림 3-10 이벤트 여부 변수 구분



예를 들어 1월 4일에 비교란에 신년 떡국 무료 배식이라는 이벤트가 표시가 되어 있는 경우 이날의 식수 인원이 전체 평균 식수 인원의 평균 보다 많은 경우 긍정 이벤트로 변수화하여 표시를 하였으며 이를 그림으로 나타내면 그림 3-11과 같다.

그림 3-11 이벤트 여부 관련 변수값 부여 예시

예시

※ 전체 식수인원 평균 = 1201 명

연번	월	일	요일	날씨	기온	충식단	인원	잔반	비고
59	1	4	월	흐림	-2~5	떡국 고추튀김 진미채 과일샐러드	1445	600	신년떡국 무료배식



긍정적인 이벤트 변수에 1 입력

부정적인 이벤트 변수에 0 입력

→ 더미변수화

6. 요일 및 계절 변수

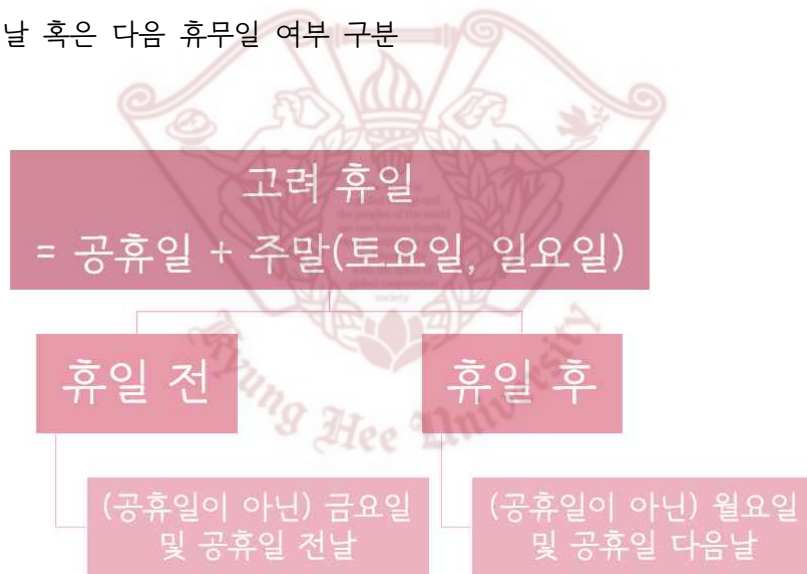
월요일부터 금요일 까지 각각 요일 별로 가변수화 하였으며, 4계절인 봄, 여름, 가을, 겨울도 계절 별로 가변수화 하였다.

7. 전날 혹은 다음날 공휴일 휴무일 여부 변수

현장 공무원과 인터뷰를 수행한 결과 통상적으로 공휴일이나 휴무일 다음 날은 전반적으로 업무 등으로 외출을 자제하는 분위기가 있어 전날 혹은 다음날이 공휴일 혹은 휴무일 여부가 식수 인원에 영향을 주는 변수로 파악되었다.

이와 관련하여 고려한 휴일은 공휴일과 주말 (토요일, 일요일)을 고려하였으며, 휴일 전 변수는 (공휴일이 아닌) 금요일 및 공휴일 전날을 의미하는 것이며, 휴일 후 변수는 (공휴일이 아닌) 월요일 및 공휴일 다음날을 의미하는 것으로 이를 그림으로 표시하면 그림 3-12와 같다.

그림 3-12 전날 혹은 다음 휴무일 여부 구분




예를 들어 2월 29일 월요일이므로 일요일인 휴일 다음날이면서 3월 1일인 삼일절 휴일의 전날에 해당되며, 이를 표시하면 그림 3-13과 같다.

그림 3-13 전날 혹은 다음 휴무일 관련 변수값 부여 예시

2016년 3월						
일	월	화	수	목	금	토
28	29	1 (삼일절)	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30	31		

 : 휴일 다음날

 : 휴일 전날

 : 휴일 다음날&전날

더미화

date	...	휴일 다음날	휴일 전날
20160229		1	1
...			
20160303		0	0
20160304		0	1
20160307		1	0

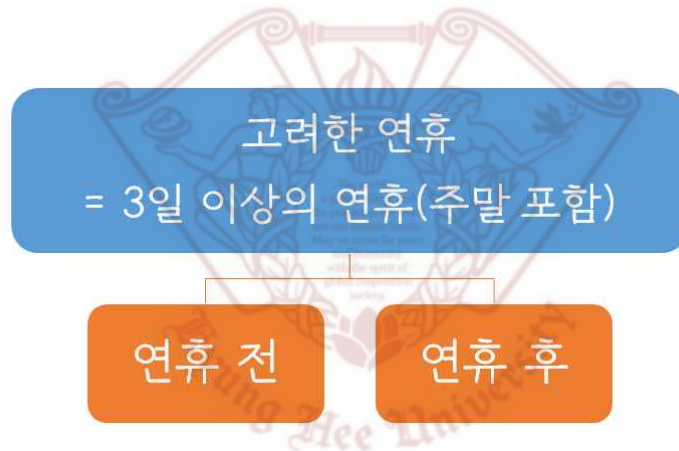


8. 전 후 연휴 및 연말 여부 변수

현장의 공무원과 인터뷰를 해본 결과 통상적으로 연휴 다음 날은 전반적으로 업무 등으로 외출을 자제하는 분위기가 있고, 연말에는 행사나 약속이 많아서 연휴 전날 혹은 다음날 여부 그리고 연말이 식수 인원에게 영향을 주는 변수로 파악되었다.

이와 관련하여 고려한 휴일은 3일 이상의 연휴(주말 포함)를 고려하였으며, 연휴 전 변수는 3일 이상의 연휴 전날을 의미하는 것이며, 연휴 후 변수는 3일 이상의 연휴 다음날을 의미하는 것이며, 이를 그림으로 표시하면 그림 3-14와 같다.

그림 3-14 연휴 전 혹은 연휴 후 여부



예를 들어 2016년 9월 14일부터 18일(주말 포함)까지 추석 연휴 기간이었었는데 이 경우 9월 13일은 연휴 전일이었으며, 9월 19일은 연휴 다음날이다. 이를 위한 구분 표시는 그림 3-15와 같다.

그림 3-15 연휴 전 혹은 연휴 후 관련 데이터 전처리 예시

2016년 9월						
일	월	화	수	목	금	토
28	29	30	31	1	2	3
4	5	6	7	8	9	10
11	12	13	14	15(추석)	16	17
18	19	20	21	22	23	24
25	26	27	28	29	30	31

: 연휴 다음날
 : 연휴 전날

2가지 변수로 분류

date	...	연휴 다음날	연휴 전날
20160901		0	0
...			
20160913		0	1
20160919		1	0

연말은 12월 21일부터 31일로 하였으며, 이를 그림으로 표시하면 그림 3-16과 같다.

그림 3-16 연말 관련 데이터 전처리

2017년 12월						
일	월	화	수	목	금	토
26	27	28	29	30	1	2
3	4	5	6	7	8	9
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	30
31						

: 연말*

date	...	연말
20171201		0
...		
20171221		1
...		...
20171229		1

제 4 장 분석 예측 모델

제 1 절 분석 예측 모델 비교

본격적인 예측 모형 구축에 앞서 다중 회귀 모델을 적용, 8개 카테고리의 63개의 모든 변수를 사용한 다중 회귀 모델링의 식이 어느 정도의 설명력을 가지는지를 파악하였다.

63개 모든 변수를 고려한 다중 회귀 모델링의 결과는 다음과 같다. 그 결과 F 검정 값은 25.9이며, 회귀식의 설명력은 66.47%가 나왔고 p-Value (p-value: $< 2.2e-16$)도 매우 낮아 유의미한 것으로 나왔다.

```
linear.3year <- lm(l.men ~., data = meal) # 모든 관측치 적용
```

Call:

```
lm(formula = l.men ~ ., data = meal)
```

Residuals:

Min	1Q	Median	3Q	Max
-566.56	-60.91	-0.54	60.06	362.23

Residual standard error: 97.42 on 681 degrees of freedom

Multiple R-squared: 0.6914, Adjusted R-squared: 0.6647

F-statistic: 25.86 on 59 and 681 DF, p-value: $< 2.2e-16$

그리고 6가지의 모델링을 진행함에 있어서 S시에서 제공한 3년치 데이터 중 Training Set과 Test Set을 7:3으로 구분, 3년치의 데이터 중 70%에 해당하는 518개의 데이터를

Training Set으로 진행 그리고 나머지 30%에 해당하는 223개의 데이터를 Test Set으로 진행을 하였다.

그리고 63개의 변수들로 이루어진 다중 회귀식의 회귀 계수는 다음과 같다.

Coefficients: (3 not defined because of singularities)

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-1.072e+03	1.198e+03	-0.895	0.371259
b.men	1.830e-01	3.139e-02	5.828	8.66e-09 ***
avail.men	1.173e-01	4.164e-02	2.818	0.004980 **
연휴전날	-4.785e+01	2.822e+01	-1.695	0.090463 .
연휴다음날	3.994e+01	2.949e+01	1.354	0.176116
복날	2.693e+02	4.385e+01	6.141	1.40e-09 ***
무료떡국	6.937e+01	6.444e+01	1.076	0.282108
대보름	1.599e+02	6.147e+01	2.601	0.009487 **
불쾌지수	1.739e+00	2.045e+00	0.850	0.395584
체감온도	-2.079e+00	2.080e+00	-0.999	0.318024
폭염	2.042e+01	1.981e+01	1.031	0.303057
rain	1.563e+01	1.814e+01	0.861	0.389388
snow	-2.000e+01	4.538e+01	-0.441	0.659646
강수량	-2.571e+00	4.500e+00	-0.571	0.567912
spring	-1.079e+01	1.492e+01	-0.723	0.469804
summer	-9.495e+00	2.141e+01	-0.443	0.657607
fall	-8.609e+00	1.551e+01	-0.555	0.579051
winter	NA	NA	NA	NA
mon	-9.645e+01	3.987e+02	-0.242	0.808915
tues	-7.913e+01	2.467e+02	-0.321	0.748515
wed	-8.856e+01	2.507e+02	-0.353	0.724013
thurs	-4.640e+01	1.796e+02	-0.258	0.796184

fri	NA	NA	NA	NA
bfr_holi	-1.591e+02	2.716e+01	-5.857	7.33e-09 ***
aft_holi	9.113e+01	2.649e+01	3.440	0.000618 ***
연말	-1.784e+02	2.712e+01	-6.578	9.53e-11 ***
곡류	-5.938e-01	8.544e+00	-0.070	0.944611
두류	-1.255e+00	8.670e+00	-0.145	0.884965
난류	2.912e+01	1.267e+01	2.298	0.021882 *
목류	-1.195e+01	1.417e+01	-0.843	0.399385
어패류	7.477e+00	7.020e+00	1.065	0.287185
육류	3.307e+01	8.075e+00	4.095	4.72e-05 ***
채소류	-7.084e+00	4.600e+00	-1.540	0.123994
해조류	-6.389e+00	8.777e+00	-0.728	0.466923
떡류	1.187e+01	1.309e+01	0.907	0.364532
양념.및.장류	1.253e+01	6.731e+00	1.862	0.063100 .
김치류	-1.200e+01	8.347e+00	-1.437	0.151166
만두류	4.387e+00	1.647e+01	0.266	0.790056
면류	7.189e+00	9.495e+00	0.757	0.449189
과일류	-1.066e+01	1.092e+01	-0.976	0.329574
밥류	7.442e+00	9.378e+00	0.794	0.427758
죽류	9.407e+01	4.153e+01	2.265	0.023811 *
덮밥국밥류	4.658e+01	2.171e+01	2.146	0.032265 *
비빔밥볶음밥류	5.873e+01	1.746e+01	3.364	0.000813 ***
김.초.밥류	NA	NA	NA	NA
국수류	1.496e+01	3.140e+01	0.476	0.634065
국탕류	-5.122e+01	2.389e+01	-2.144	0.032376 *
찌개류	-2.285e+01	2.562e+01	-0.892	0.372853
구이류	3.366e+01	1.227e+01	2.744	0.006230 **
무침류	5.898e+00	9.176e+00	0.643	0.520565
볶음류	-2.783e+00	1.078e+01	-0.258	0.796333

장아찌류	-2.503e+01	1.930e+01	-1.297	0.195156
전류	2.975e+01	1.350e+01	2.204	0.027838 *
조림류	1.279e+01	1.191e+01	1.074	0.283228
찜류	7.414e+00	1.505e+01	0.493	0.622387
튀김류	2.475e+01	1.108e+01	2.233	0.025841 *
샐러드류	1.706e+01	1.279e+01	1.334	0.182704
단품류	-5.628e+00	6.948e+00	-0.810	0.418203
유제품	-1.293e+01	2.096e+01	-0.617	0.537356
빵과자류	-4.247e+01	2.576e+01	-1.649	0.099670 .
음료.및.주류	-2.551e+00	2.246e+01	-0.114	0.909591
요일평균	1.065e+00	1.187e+00	0.897	0.369863
월평균	4.980e-01	8.987e-02	5.542	4.29e-08 ***

또한 실제값과 예측 값을 가지고 오차율을 구하였으며, 오차율을 구하는 식은 다음과 같다.

$$\text{오차율} = \frac{\text{절대값}(\text{실제값} - \text{예측값})}{\text{실제값}}$$

또한 Training Set을 추출하는 데 있어 비복원 추출을 활용하여 진행을 하였으며, 샘플링 오류를 줄이기 위해 무작위로 샘플링을 11번을 진행하여 후진 단계적 선택법, LASSO, Ridge, Random Forest, Bagging, Boosting 등의 6가지 모델링의 예측 값을 구하여 각각 개별의 예측치와 실제값을 가지고 오차율을 비교 분석하였으며, 오차율을 최소화 하는 최종 모델링을 진행하였다.

1. 후진 단계적 선택법 (Backward Stepwise Selection)

다중 선형 회귀란 종속 변수 Y 와 두 개 이상의 독립변수 X 와의 관계를 설명하는 모델링 기법이다.

$$y_i = \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i$$

다중 선형 회귀 모델 중 후진 단계적 선택법 (Backward Stepwise Selection)은 모든 변수들을 다 포함한 상태에서 유의성이 낮은 변수들을 제외시키는 변수 선택법이다.

Backward Stepwise Selection 의 필요성 및 장점은 다음과 같다.

- 유의성과 설명력이 높은 간단한 모델을 형성하고자 할 때 필요하다.
- 변수의 수가 많아서 유의성이 높은 변수를 선택하기 어려울 때 유용하다.

223개의 데이터를 Test Set으로 Backward Stepwise Regression 모델링 결과는 다음과 같다. 전체 63개의 설명 변수 중 27개의 변수를 가지고 다중 회귀식이 만들어 졌으며, 이 회귀식의 F 검정 값은 37.4이며, 회귀식의 설명력은 65.51%가 나왔고 p 값(p-value: $< 2.2e-16$)도 매우 낮아 유의미한 것으로 나왔다.

Call:

```
lm(formula = l.men ~ b.men + avail.men + 연휴다음날 + 복날 + 대보름 + mon +  
tues + wed + thurs + bfr_holi + aft_holi + 연말 + 난류 + 육류 + 떡류 + 면류 +  
죽류 + 덮밥국밥류 + 비빔밥볶음밥류 + 국탕류 + 구이류 + 볶음류 + 장아찌류 + 전류  
+ 찜류 + 튀김류 + 월평균, data = meal.train)
```

Residuals:

Min	1Q	Median	3Q	Max
-----	----	--------	----	-----

-265.81 -58.99 -1.65 60.36 351.50

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-6.20237	119.99691	-0.052	0.958799
b.men	0.18989	0.03599	5.277	1.98e-07 ***
avail.men	0.08676	0.04034	2.151	0.031958 *
연휴다음날	91.99159	33.49604	2.746	0.006248 **
복날	213.12146	49.19736	4.332	1.79e-05 ***
대보름	176.93421	97.86628	1.808	0.071232 .
mon	282.10796	23.94849	11.780	< 2e-16 ***
tues	164.38859	19.97824	8.228	1.73e-15 ***
wed	149.77784	18.66479	8.025	7.59e-15 ***
thurs	141.05234	19.20350	7.345	8.64e-13 ***
bfr_holi	-164.90235	30.60582	-5.388	1.11e-07 ***
aft_holi	93.91827	30.60068	3.069	0.002266 **
연말	-159.80488	31.11780	-5.135	4.07e-07 ***
난류	22.48698	13.73065	1.638	0.102122
육류	32.03839	8.55283	3.746	0.000201 ***
떡류	25.44905	13.78777	1.846	0.065529 .
면류	17.82913	9.83477	1.813	0.070464 .
죽류	101.91024	52.69851	1.934	0.053709 .
덮밥국밥류	53.22915	23.18755	2.296	0.022121 *
비빔밥볶음밥류	32.06270	17.82649	1.799	0.072698 .
국탕류	-39.15747	10.61595	-3.689	0.000251 ***
구이류	32.57798	11.49766	2.833	0.004795 **
볶음류	-17.11803	9.05155	-1.891	0.059191 .
장아찌류	-41.51043	21.02769	-1.974	0.048933 *
전류	35.60047	12.97786	2.743	0.006308 **

점류 -27.97258 16.39402 -1.706 0.088592 .
 튀김류 30.68442 10.95672 2.801 0.005304 **
 월평균 0.56145 0.09644 5.822 1.05e-08 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 96.17 on 490 degrees of freedom

Multiple R-squared: 0.6731, Adjusted R-squared: 0.6551

F-statistic: 37.36 on 27 and 490 DF, p-value: < 2.2e-16

보다 정확한 예측치를 위해 S시에서 제공한 3년치 데이터 중 Training Set과 Test Set을 7:3으로 구분, 3년치의 데이터 중 70%에 해당하는 518개의 데이터를 Training Set으로 진행 그리고 나머지 30%에 해당하는 223개의 데이터를 Test Set으로 구분하여, 각각 다르게 11번의 Sampling을 진행을 하였고, 이에 따른 223개의 test set에 의한 예측치와 실제값의 오차율을 계산하였으며, 이 회귀식을 통한 식수 예측 인원과 실제 식사 인원의 오차율은 표 4-1과 같다.

표 4-1 각 seed 별 Backward Stepwise Selection에 의한 예측값과 실제값의 오차율

모델링	seed 1	seed 2	seed 3	seed 4	seed 5	seed 6	seed 7	seed 8	seed 9	seed 10	seed 123	평 균
back ward 오차율	7.55%	7.46%	6.97%	7.71%	7.42%	6.82%	7.00%	7.08%	7.14%	7.45%	6.68%	7.21 %

표 4-1에서 보듯이 S시에서 제공한 3년치 데이터 중 Training Set과 Test Set을 7:3으로 구분, 3년치의 데이터 중 70%에 해당하는 518개의 데이터를 Training Set으로 진행하고 그리고 나머지 30%에 해당하는 223개의 데이터를 Test Set으로 구분하여 Sampling시 seed를 어떻게 하느냐에 따라 오차율이 최소 6.68%에서 최대 7.55% 까지 약 0.87%의 차이가 난다는 것을 알 수 있으며, 이는 Sampling에 따라 오차율이 크게

변동을 가져 올 수 있다는 것을 확인시켜 주었다.

11번의 각각 sampling을 통해 후진 단계적 선택법에 의한 식수 예측 모델링의 오차율의 평균값은 7.21%이었다.

또한 후진 단계적 선택법에 의한 식수 예측 모델링 적용시 사용된 설명 변수들간의 다중 공선성을 알아 보기 위해 분산 팽창 계수를 구하였으며 그 결과는 표 4-2와 같다.

통상적으로 분산 팽창 계수가 8이상인 경우 다중 공선성이 있는 것으로 판단하는데, 표 4-2에서 보듯이 후진 단계적 선택법에 의한 식수 예측 모델링 적용시 사용된 설명 변수들 간에 다중 공선성은 없는 것으로 파악되었다.

표 4-2 각 변수 별 분산 팽창 계수

변수	b.men	avail .men	복날	무료 떡국	대보 름	강수 량	bfr_h oli	aft_h oli	연말	난류	어패 류
VIF 값	2.06	1.32	1.05	1.15	1.06	1.11	1.05	1.08	1.37	1.08	1.24
변수	육류	양념.및.장류		죽류	덮밥국밥류		비빔밥볶음밥 류		국탕 류	구이 류	전류
VIF 값	1.53	1.22		1.19	1.14		1.11		1.18	1.19	1.15
변수	튀김 류	단품 류	빵과 자류	요일 평균	월평 균						
VIF 값	1.29	1.17	1.08	3.06	1.48						

2. LASSO

LASSO란 기존의 회귀분석에 제약조건(t)을 주어 중요하지 않은 변수의 회귀계수 값을 축소시켜 0으로 만들어 주는 모델이다.

$$\hat{\beta}^{lasso} = \min_{\beta} \sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^p x_{ij}\beta_j)^2$$

$$\sum_{j=1}^p |\beta_j| \leq t$$

LASSO의 필요성 및 장점은 다음과 같다.

- 변수가 많을 경우, 간결함의 원리에 따라 가장 잘 설명할 수 있는 변수만을 골라내어 간단한 설명으로 만들어야 할 필요가 있다.
- 영향력이 적은 변수의 계수를 0으로 만듦으로써 어떠한 변수가 모델에 중요한지 알게 되어 모델 해석력이 좋아진다.

223개의 데이터를 Test Set으로 LASSO 회귀의 모델링을 한 결과는 표 4-3과 같다.
(계수가 0이 아닌 것은 45개임)

표 4-3 LASSO 회귀 모델링 결과의 변수들의 계수값

번호	변수	계수	번호	변수	계수	번호	변수	계수
1	b.men	0.189	16	thurs	0.101	31	죽류	32.2
2	avail.men	0.123	17	bfr_holi	-140	32	덮밥국밥류	12.3
3	연휴전날	-41.7	18	aft_holi	70.7	33	비빔밥 볶음밥류	31.4
4	연휴다음날	30.1	19	연말	-157	34	국수류	15.9
5	복날	217	20	난류	20.2	35	국탕류	-25.9
6	무료떡국	63	21	목류	-8.94	36	구이류	21.6
7	대보름	103	22	어패류	3.87	37	볶음류	-5.73
8	체감온도	-0.0037	23	육류	27.6	38	장아찌류	-24.4

9	폭염	6.65	24	채소류	-6.15	39	전류	12.1
10	rain	5.25	25	해조류	-0.062	40	튀김류	10.6
11	snow	-0.323	26	떡류	0.277	41	샐러드류	5.46
12	spring	-1.04	27	양념및장류	6.98	42	단품류	-6.63
13	winter	5.28	28	김치류	-5.56	43	빵과자류	-33.3
14	mon	20.1	29	면류	1.20	44	요일평균	0.648
15	wed	-0.0044	30	과일류	-0.117	45	월평균	0.486

S시에서 제공한 3년치 데이터 중 Training Set과 Test Set을 7:3으로 구분, 3년치의 데이터 중 70%에 해당하는 518개의 데이터를 Training Set으로 분석을 진행하였다.

그리고 나머지 30%에 해당하는 223개의 데이터를 Test Set으로 구분하여 각각 다르게 11번의 Sampling을 진행을 하였으며, 이에 따른 223개의 test set에 의한 예측치와 실제값의 오차율을 계산하였다.

LASSO 모델링을 통한 식수 예측 인원과 실제 식사 인원의 오차율은 표 4-4와 같다. 표 4-4에서 보듯이 S시에서 제공한 3년치 데이터 중 Training Set과 Test Set을 7:3으로 구분, 3년치의 데이터 중 70%에 해당하는 518개의 데이터를 Training Set으로 진행하고 나머지 30%에 해당하는 223개의 데이터를 Test Set으로 구분하여 Sampling시 seed를 어떻게 하느냐에 따라 오차율이 최소 6.56%에서 최대 7.59% 까지 약 1.03%의 차이가 난다는 것을 알 수 있으며, 이는 후진 단계적 선택법과 마찬가지로 Sampling에 따라 오차율이 크게 변동을 가져 올 수 있다는 것을 확인시켜 주었다.

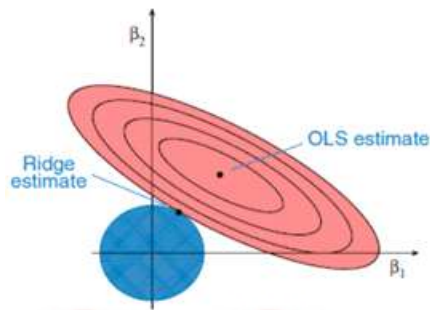
11번의 각각 sampling을 통해 LASSO에 의한 식수 예측 모델링의 오차율의 평균값은 7.14%이었다.

표 4-4 각 seed 별 LASSO에 의한 예측값과 실제값의 오차율

모델링	seed 1	seed 2	seed 3	seed 4	seed 5	seed 6	seed 7	seed 8	seed 9	seed 10	seed 123	평균
LASSO 오차율	7.59 %	7.28 %	7.15 %	7.56 %	7.33 %	6.98 %	6.74 %	6.87 %	7.06 %	7.39 %	6.56 %	7.14 %

3. Ridge Regression

RIDGE regression은 기존의 회귀분석에 제약조건(t^2)을 주어 중요하지 않은 변수의 회귀 계수 값을 축소시켜 0에 가깝게 만들어 주는 모델이다.



RIDGE regression의 특징은 다음과 같다.

- 변수가 많을 경우, 오차를 최소로 하면서 최대한 간단한 모델로 만들 수 있다.
- 적합이 되는 해에 제약조건을 부여함으로써 모델의 과적합을 방지할 수 있다.

223개의 데이터를 Test Set으로 Ridge 회귀의 모델링을 한 결과는 표 4-5와 같다.
(상위 30개 만 기술함)

표 4-5 Ridge 회귀 모델링 결과의 변수들의 계수값

번호	변수	계수	번호	변수	계수	번호	변수	계수
1	복날	177.99	11	비빔밥	20.16	21	summer	8.58
2	대보름	124.95	12	볶음밥류	19.46	22	wed	7.47
3	무료떡국	84.74	13	전류	15.92	23	찌개류	5.68
4	aft_holi	71.13	14	덮밥국밥류	14.21	24	유제품	4.24
5	mon	62.89	15	샐러드류	14.02	25	떡류	3.3
6	죽류	54.51	16	강수량	11.64	26	무침류	2.66
7	국수류	29.65	17	폭염	10.8	27	밥류	2.47

8	구이류	27.32	18	어패류	9.99	28	tues	1.36
9	연휴다음날	26.36	19	양념및장류	9.99	29	fail	1.12
10	육류	24.57	20	튀김류	8.9	30	월평균	0.42

S시에서 제공한 3년치 데이터 중 Training Set과 Test Set을 7:3으로 구분, 3년치의 데이터 중 70%에 해당하는 518개의 데이터를 Training Set으로 진행 그리고 나머지 30%에 해당하는 223개의 데이터를 Test Set으로 구분하여 각각 다르게 11번 샘플링을 진행을 하였으며, 이에 따른 223개의 test set에 의한 예측치와 실제값의 오차율을 계산하였다.

Ridge 모델링을 통한 식수 예측 인원과 실제 식사 인원의 오차율은 표 4-6과 같다.

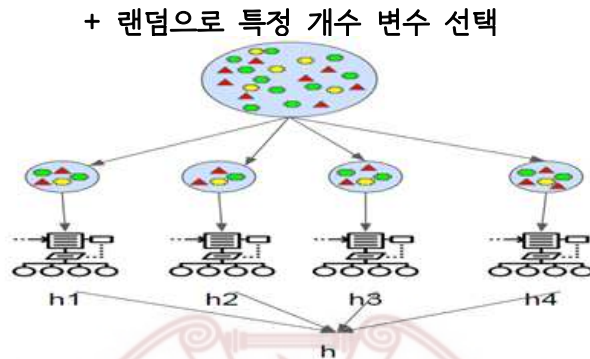
표 4-6 각 seed 별 Ridge에 의한 예측값과 실제값의 오차율

모델링	seed 1	seed 2	seed 3	seed 4	seed 5	seed 6	seed 7	seed 8	seed 9	seed 10	seed 123	평균
Ridge 오차율	7.63 %	7.23 %	7.18 %	7.61 %	7.40 %	6.97 %	6.69 %	6.98 %	7.20 %	7.32 %	6.57 %	7.16 %

표 4-6에서 보듯이 S시에서 제공한 3년치 데이터 중 Training Set과 Test Set을 7:3으로 구분, 3년치의 데이터 중 70%에 해당하는 518개의 데이터를 Training Set으로 진행하고 나머지 30%에 해당하는 223개의 데이터를 Test Set으로 구분하여 Sampling 시 seed를 어떻게 하느냐에 따라 오차율이 최소 6.57%에서 최대 7.63% 까지 약 1.06%의 차이가 난다는 것을 알 수 있으며, 이는 후진 단계적 선택법이나 LASSO와 마찬가지로 표본 추출에 따라 오차율이 크게 변동을 가져 올 수 있다는 것을 확인시켜 주었다.

11번의 각각 sampling을 통해 Ridge에 의한 식수 인원 예측 모델링의 오차율의 평균값은 7.16%이었다.

Random Forest란 bagging에서의 model을 의사결정나무로 이용 + 랜덤으로 특정 개수의 변수를 선택해주는 모델링 기법이다.



Random Forest 의 특징은 다음과 같다.

- 데이터를 여러 데이터 묶음들로 분할시키면서 과적합 될 위험을 막아준다
- 데이터가 많을수록 성능이 좋아진다.
- 변수를 랜덤하게 뽑음으로써 비슷한 모델들이 만들어지지 않도록 한다.

223개의 데이터를 Test Set으로 Random Forest 모델링을 한 결과는 다음과 같다.

Call:

```
randomForest(x = meal.train, y = meal.train.label, xtest = meal.test, ytest =
meal.test.label, ntree = 2000)
```

Type of random forest: regression

Number of trees: 2000

No. of variables tried at each split: 20

Mean of squared residuals: 11657.16

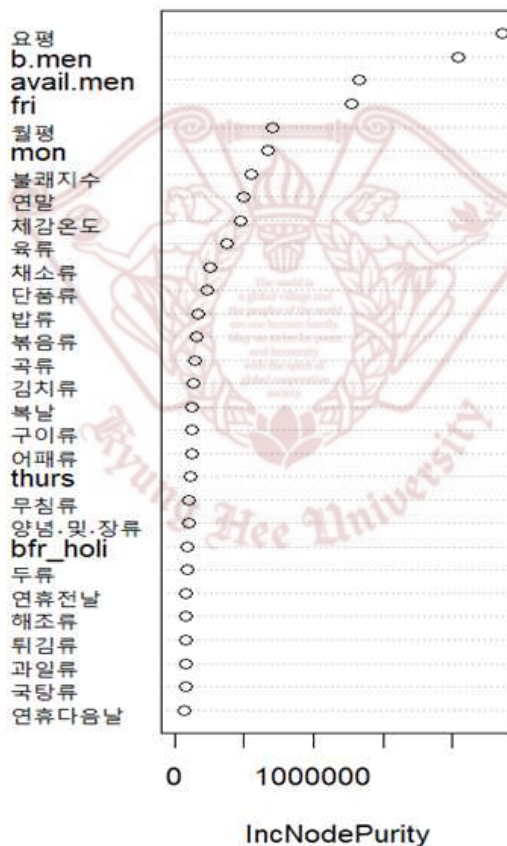
% Var explained: 58.25

Test set MSE: 11269.55

% Var explained: 61.1

Random Forest에 의한 변수별 중요도를 그림으로 나타내면 그림 4-1과 같다.

그림 4-1 Random Forest에 의한 변수별 중요도



S시에서 제공한 3년치 데이터 중 Training Set과 Test Set을 7:3으로 구분, 3년치의 데이터 중 70%에 해당하는 518개의 데이터를 Training Set으로 진행 그리고 나머지

30%에 해당하는 223개의 데이터를 Test Set으로 구분하여 각각 다르게 11번의 샘플링을 진행하였으며, 이에 따른 223개의 test set에 의한 예측치와 실제값의 오차율을 계산하였다.

Random Forest 모델링을 통한 식수 예측 인원과 실제 식사 인원의 오차율은 표 4-7과 같다.

표 4-7 각 seed 별 Random Forest에 의한 예측값과 실제값의 오차율

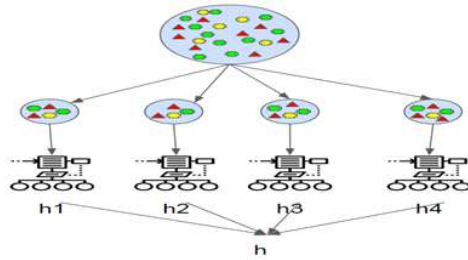
모델링	seed 1	seed 2	seed 3	seed 4	seed 5	seed 6	seed 7	seed 8	seed 9	seed 10	seed 123	평균
RF 오차율	7.65 %	7.34 %	7.58 %	7.21 %	7.57 %	7.22 %	7.17 %	7.22 %	7.57 %	7.70 %	6.45 %	7.33 %

표 4-7에서 보듯이 S시에서 제공한 3년치 데이터 중 Training Set과 Test Set을 7:3으로 구분, 3년치의 데이터 중 70%에 해당하는 518개의 데이터를 Training Set으로 진행 그리고 나머지 30%에 해당하는 223개의 데이터를 Test Set으로 구분하여 표본 추출시 seed를 어떻게 하느냐에 따라 오차율이 최소 6.45%에서 최대 7.70% 까지 약 1.25%의 차이가 난다는 것을 알 수 있으며, 이는 후진 단계적 선택법이나 LASSO, Ridge와 마찬가지로 Sampling에 따라 오차율이 크게 변동을 가져 올 수 있다는 것을 확인시켜 주었다.

11번의 각각 sampling을 통해 Random Forest에 의한 식수 예측 모델링의 오차율 평균 값은 7.33%이었다.

5. Bagging

Bagging이란 원본 훈련 데이터를 복원 추출하여 여러 데이터 묶음으로 만들고 각각 따로 모델링하여 모델들의 평균으로 결과를 내는 모델링 기법이다.



Bagging의 특징은 다음과 같다.

- 데이터를 여러 데이터 묶음들로 분할시키면서 과적합 될 위험을 막아준다.
- 데이터가 많을수록 성능이 좋아진다.

223개의 데이터를 Test Set으로 Bagging 모델링을 한 결과는 다음과 같다.

Bagging regression trees with 100 bootstrap replications

Call: bagging.data.frame(formula = l.men ~ ., data = meal.train, nbagg = 100)

전체 설명 변수 중 57개의 변수가 영향력을 지니고 있다.

S사에서 제공한 3년치 데이터 중 Training Set과 Test Set을 7:3으로 구분하고 3년치의 데이터 중 70%에 해당하는 518개의 데이터를 Training Set으로 진행하였다. 그리고 나머지 30%에 해당하는 223개의 데이터를 Test Set으로 구분하여 각각 다르게 11번의 Sampling을 진행을 하였으며, 이에 따른 223개의 test set에 의한 예측치와 실제값의

오차율을 계산하였다.

Bagging 모델링을 통한 식수 예측 인원과 실제 식사 인원의 각각의 오차율은 표 4-8과 같다.

표 4-8 각 seed 별 Bagging에 의한 예측값과 실제값의 오차율

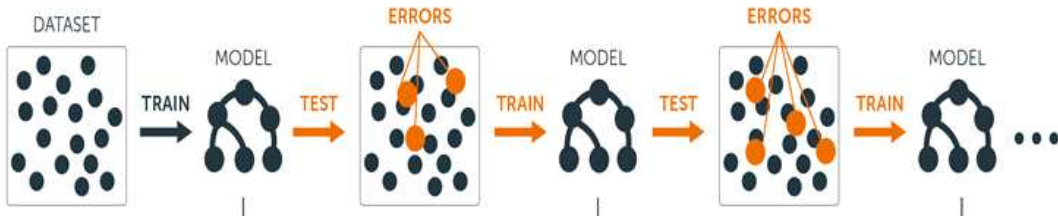
모델링	seed 1	seed 2	seed 3	seed 4	seed 5	seed 6	seed 7	seed 8	seed 9	seed 10	seed 123	평균
배깅 오차율	8.12 %	7.77 %	8.01 %	7.49 %	8.20 %	7.85 %	7.78 %	7.82%	8.03 %	8.06 %	6.86 %	7.82 %

표 4-8에서 보듯이 S시에서 제공한 3년치 데이터 중 Training Set과 Test Set을 7:3으로 구분, 3년치의 데이터 중 70%에 해당하는 518개의 데이터를 Training Set으로 진행하고 나머지 30%에 해당하는 223개의 데이터를 Test Set으로 구분하여 표본 추출 시 seed를 어떻게 하느냐에 따라 오차율이 최소 6.86% 에서 최대 8.12% 까지 약 1.26%의 차이가 난다는 것을 알 수 있으며, 이는 후진 단계적 선택법이나 LASSO, Ridge 그리고 Random Forest 등과 마찬가지로 Sampling에 따라 오차율이 크게 변동을 가져올 수 있다는 것을 확인시켜 주었다.

11번의 각각 sampling을 통해 Bagging에 의한 식수 예측 모델링의 오차율의 평균값은 7.82%이었다.

6. Boosting

Boosting이란 예측이 잘 되지 않은 데이터가 더 많이 뽑히도록 가중치를 부여하여 잘 학습되도록 하는 기법이다.



Boosting의 특징은 다음과 같다.

- 이전 모델의 오류를 고려해주기 때문에 보다 더 성능이 좋아진다.

223개의 데이터를 Test Set으로 Bagging 모델링을 한 결과는 다음과 같다.

```
gbm(formula = l.men ~ ., distribution = "gaussian", data = meal.train)
```

A gradient boosted model with gaussian loss function.

100 iterations were performed.

There were 62 predictors of which 3 had non-zero influence.

전체 62개의 설명 변수 중 3개의 변수가 영향력을 지니고 있으며, 그 결과는 표 4-9와 같다.

표 4-9 Boosting에 의한 영향력 있는 변수의 상대적 영향력

변수	rel.inf
요일 평균	43.848
fri	34.107
b,men	21.945

S사에서 제공한 3년치 데이터 중 Training Set과 Test Set을 7:3으로 구분하였으며, 3년치의 데이터 중 70%에 해당하는 518개의 데이터를 Training Set으로 진행 그리고 나머지 30%에 해당하는 223개의 데이터를 Test Set으로 구분하여 각각 다르게 11번의 표본 추출을 진행하였으며, 이에 따른 223개의 test set에 의한 예측치와 실제값의 오차율을 계산하였다.

Boosting 모델링을 통한 식수 예측 인원과 실제 식사 인원의 각각의 오차율은 표 4-10과 같다.

표 4-10 각 seed 별 Boosting에 의한 예측값과 실제값의 오차율

모델링	seed 1	seed 2	seed 3	seed 4	seed 5	seed 6	seed 7	seed 8	seed 9	seed 10	seed 123	평균
boosting 오차율	12.1 %	11.3 %	11.8 %	12.0 %	12.9 %	11.8 %	11.5 %	11.6 %	12.0%	13.4 %	11.0 %	11.9 %

표 4-10에서 보듯이 S사에서 제공한 3년치 데이터 중 Training Set과 Test Set을 7:3으로 구분, 3년치의 데이터 중 70%에 해당하는 518개의 데이터를 Training Set으로 진행하고 그리고 나머지 30%에 해당하는 223개의 데이터를 Test Set으로 구분하여 표본 추출시 seed를 어떻게 하느냐에 따라 오차율이 최소 11.0%에서 최대 13.4% 까지 약 2.4%의 차이가 난다는 것을 알 수 있으며, 이는 후진 단계적 선택법이나 LASSO, Ridge, Random Forest, Bagging 등과 마찬가지로 Sampling에 따라 오차율이 크게 변동을 가져 올 수 있다는 것을 확인시켜 주었다.

11번의 각각 sampling을 통해 Boosting에 의한 식수 예측 모델링의 오차율의 평균값은 11.9%이었으며, 또한 오차율의 값이 너무 커서 Boosting 모델링으로 거의 의미가 없는 방법이었다.

제 2 절 분석 예측 모델 결과 평가

후진 단계적 선택법, LASSO, Ridge, Random Forest, Bagging 그리고 Boosting 등 6가지 모델링 기법을 적용하여 실제 식수 인원과 각각 모델링 기법 예측치의 오차율을 정리하면 표 4-11과 같다.

표 4-11 각 seed 별 모델링 기법에 의한 예측값과 실제값의 오차율 비교

모델링	seed 1	seed 2	seed 3	seed 4	seed 5	seed 6	seed 7	seed 8	seed 9	seed 10	seed 123	평균
back ward 오차율	<u>7.55</u> %	7.46 %	<u>6.97</u> %	7.71 %	7.42 %	<u>6.82</u> %	7.00 %	7.08 %	7.14 %	7.45 %	6.68%	7.21 %
LASSO 오차율	7.59 %	7.28 %	7.15 %	7.56 %	<u>7.33</u> %	6.98 %	6.74 %	<u>6.87</u> %	<u>7.06</u> %	7.39 %	6.56%	7.14 %
Ridge 오차율	7.63 %	<u>7.23</u> %	7.18 %	7.61 %	7.40 %	6.97 %	<u>6.69</u> %	6.98 %	7.20 %	<u>7.32</u> %	6.57%	7.16 %
RF 오차율	7.65 %	7.34 %	7.58 %	<u>7.21</u> %	7.57 %	7.22 %	7.17 %	7.22 %	7.57 %	7.70 %	<u>6.45</u> %	7.33 %
bagging 오차율	8.12 %	7.77 %	8.01 %	7.49 %	8.20 %	7.85 %	7.78 %	7.82 %	8.03 %	8.06 %	6.86%	7.82 %
boosting 오차율	12.1 %	11.3 %	11.8 %	12.0 %	12.9 %	11.8 %	11.5 %	11.6 %	12.0 %	13.4 %	11.0%	11.9 %

표 4-11에서 보듯이 Bagging과 Boosting을 제외한 상위 4개 모델링인 Backward, Ridge, LASSO, Random Forest 모델링은 seed에 따라 각 모델마다 오차율이 낮은 특정 데이터 분포가 있는 것을 확인할 수 있었다.

예를 들어 Sampling시 seed 4와 123으로 진행을 한 경우 오차율이 제일 낮은 것은 Random Forest이었으며, seed가 1, 3 그리고 6인 경우 오차율이 제일 낮은 모델은 후진 단계적 회귀 모형이었으며, seed 5, 8 그리고 9로 진행 한 경우는 LASSO가 제일 오차율이 작았으며, seed 2, 7 그리고 10인 경우는 Ridge가 제일 오차율이 작은 것을 확인할 수 있었다.

이는 어떤 특정 모델이 오차율이 제일 낮은 최상의 모델링을 생성한다는 것이 아니라는 것을 의미한다.

본 연구를 통해 사용되는 데이터에 따라 적용된 각 모델링의 오차율이 낮은 정도가 달라, 향후 미래의 데이터를 가지고 식수 예측을 진행하는 경우 최적의 모델링을 생성하기 위해 하나의 기법에 따른 특정 모델링을 적용하기에 무리가 따른다는 것을 알 수 있었다.



제 3 절 분석 예측 최종 모델링

하나의 기법에 의해 특정 모델링을 진행하는 경우 미래의 데이터에 대한 예측시 오차율이 커질 수 있는 경우가 발생할 수 있어 주의가 요구되며, 1차 모델링의 결과를 기본으로 각각 개발된 모델에 대한 예측력을 비교 평가하고 평가된 모델을 기반으로 향후 미래의 데이터 예측의 안정성을 위해 예측력이 높은 최적의 모델을 생성하였다.

최종 모델링은 다음의 2단계를 거쳐 진행을 하였다.

단계1) 1차 모델링을 적용한 기법 중 오차율 상위 4개 선정

단계2) 오차율 상위 4개 모델링의 예측치 값들의 평균값을 최종 모델링의 예측 값으로 선정

우선적으로 오차율 상위 4개 모델의 예측치 평균은 표 4-12와 같다.

표4-12 상위 4개 모델링의 오차율 평균

오차율	LASSO	Ridge	Backward	Random Forest
평균	7.14%	7.16%	7.21%	7.33%

예측 모델의 안정화 및 최상의 모델링을 위해 진행한 최종 모델링은 오차율 상위 4개 모델의 예측치의 평균값을 구해서 이를 최종 모델링의 예측값으로 진행하였으며 이를 표로 나타내면 표 4-13과 같다.

표 4-13 최종 모델링 방법

LASSO 예측치	+	Ridge 예측치	+	Backward 예측치	+	RF 예측치	=	최종 모델링	실제 식수인원
830		819		833		850		833	766

1183		1174		1197		1146		1175	1136
940		939		938		880		924	929
1146		1145		1145		1192		1157	1185
---		---		---		---		---	---

최종 모델링은 하나의 기법에 의해 특정 모델링을 진행하는 경우 미래의 데이터에 대한 예측시 오차율이 커질 수 있는 경우가 발생할 수 있어 1차 모델링의 결과를 기본으로 모델별로 특정의 Sampling에 강한 것이 있는 것을 활용한 것이다.

각각 개발된 모델에 대한 예측력을 비교 평가하고 비교 평가된 모델을 기반으로 향후 미래의 데이터 예측의 안정성을 위해 예측력이 높은 최적의 모델을 생성 진행하였다.

상위 4개의 각각의 모델링의 예측치들의 평균값을 최종 모델링의 예측값으로 진행을 한 결과 오차율은 상위 4개 모델링의 오차율의 각각의 평균 (LASSO 7.14%, Ridge 7.16%, Backward 7.16%, 그리고 Random Forest 7.33%) 최소값이 7.14 보다 낮은 7.04의 최저의 오차율의 평균값을 가지면서 특정 Sampling에 따라 예측치의 값이 편중되지 않는 안정적인 예측치를 가지게 되었다.

이로서 1차 모델링을 거쳐 각각 개발된 모델에 대한 예측력을 비교 평가하고 비교평가된 모델을 기반으로 향후 미래의 데이터 예측의 안정성을 위해 예측력이 높은 최적의 모델을 생성 진행함으로써 오차율 7% 수준의 식수 예측 모델링 개발을 완료하게 되었다.

본 연구에서 1차 모델링과 최종 모델링에 의한 식수 예측 모델링은 특정 데이터셋에 과적합 현상을 피할 수 있는 방법으로 진행함으로써 미래의 실제 데이터를 기반으로 식수 인원 예측을 진행할 때 유사한 결과를 얻을 수 있다는 점에는 의미가 있다고 본다.

제 4 절 메뉴 개수 반영에 따른 식수 예측 모델링 비교

S시 집단 급식소의 3년치 (2015년 5월 1일부터 2018년 4월 30일 까지) 식단 메뉴를 예시를 하면 표 4-14과 같은 유형으로 식단이 제공이 되고 있다.

표 4-14 3년간의 식단 메뉴 예시

2015년 12월1일(화)	2015년 12월16일(수)	2016년 8월31일(수)	2015년 5월18일(월)	2015년 7월13일(월)	2017년 7월16일(금)
육개장	해물날치알 볶음밥	잔치국수	보리밥/쌀밥	닭다리백숙	나물비빔밥 &고추장
가자미구이	얼큰어묵국	닭다리바베큐	북어채무국	오징어 실채볶음	무맑은국
어묵조림	함박스테이크	고추지무침	순대야채볶음	오이고추 쌈장무침	고구마맛탕
오이생채	쫄면 무침	후식과일	고구마샐러드	포기김치	오이지무침
	양상추샐러드		상추겉절이	아이스크림	포기김치
			포기김치		굴

표 4-14 처럼 S시의 3년치 식단 메뉴를 분석을 해본 결과 S시의 제공 메뉴는 크게 4가지 유형으로 표 4-15와 같이 분류될 수 있다.

표 4-15 메뉴 유형

Case 1	Case 2	Case 3	Case 4
국류, 면류, 찌개류	비빔밥볶음밥류	밥류	떡국/백숙 등 특식
주메뉴	국류찌개류	국/찌개류/면류	메뉴1
메뉴1	주메뉴	주메뉴	메뉴2
메뉴2	메뉴1	메뉴1	메뉴3
메뉴3	메뉴2	메뉴2	메뉴4
메뉴4	메뉴3	메뉴3	메뉴5

사람들이 식단을 보고 식사 여부를 결정을 하는 경우 대체로 주메뉴를 보고 의사 결정을 하는 것으로 가정하면 위의 유형에서 보듯이 주메뉴가 메뉴에서 차지하는 순서는 Case4와 같이 처음인 경우도 있으나 Case1과 같이 2번째 그리고 Case2와 3과 같이 메뉴에서 3번째의 순서를 차지하는 경우가 있어 메뉴에 따라 주메뉴가 차지하는 순서가 다른 것을 알 수 있다.

그러므로 식수 예측 모델링을 진행시 메뉴의 중요도를 모델링에 직접 반영하고자 식단 분석 결과를 바탕으로 주메뉴를 포함하여 3가지의 메뉴 모델링에 고려하는 경우와 모든 메뉴를 모델링에 반영하는 경우를 비교하여 모델링의 결과를 분석하여 진행을 한 결과는 표 4-16과 같다.

표 4-16 메뉴 3개와 메뉴 전체에 대한 모델링의 오차율 결과

모델링	주메뉴 포함한 메뉴 3개 반영 오차율	메뉴 전체 반영 오차율
Backward Stepwise Selection	7.13%	7.21%
Ridge	7.20%	7.16%
LASSO	7.19%	7.14%
Random Forest	7.36%	7.33%
bagging	8.04%	7.82%
boosting	12.0%	11.9%

표 4-16의 결과를 통해 메뉴 3개와 메뉴 전체에 대한 모델링 진행을 한 결과, 두 가지의 오차율이 매우 유사하다는 것을 알 수 있다. 또한 메뉴 3개를 반영하였을 때 모델의 오차율이 더 낮은 경우도 발생하였으나, 전반적으로 메뉴 전체를 반영한 모델의 오차율이 근소하게 더 낮았다.

또한 주메뉴 포함한 메뉴 3개를 반영하는 모델링과 메뉴 전체를 반영하는 경우의 모델링 진행을 한 결과는 표 4-17과 같다.

표 4-17 메뉴 개수에 따른 최종 모델링의 오차율 결과

	주메뉴 포함한 메뉴 3개 반영 오차율	메뉴 전체 반영 오차율
최종 모델링	7.13%	7.04%

메뉴 개수에 따른 모델링을 진행한 결과 주메뉴를 포함한 메뉴 3개를 반영한 모델링 보다 표 4-17에서 보듯이 메뉴 전체를 반영한 모델링의 오차율이 근소하게 더 작았음을 알 수 있었다. 향후 식수 인원 예측 모델링을 진행하는 데 있어서 주메뉴 위주의 예측과 메뉴 전체의 모델링을 진행하는 것은 큰 차이가 없어 보이나, 본 연구에서는 오차율을 더 줄이기 위해 메뉴 전체를 반영한 모델링을 적용하여 진행하였다.

제 5 절 예측 최종 모델링 적용시 예상 효과

본 연구의 식수 인원 예측 알고리즘 모델링의 결과로 예상되는 직접적인 효과는 경험치의 의한 평균 식수 인원 예측 오차율이 10% ~ 11%대에서 오차율이 약 7% 대로 줄어드는 효과일 것으로 예상하며, 이로 인해 약 잔반량이 약 40% 정도 감축될 수 있을 것으로 예상된다.

이를 금액으로 환산하면, 음식물 쓰레기 절감으로 인한 쓰레기 처리 비용과 식재료 구매 비용 절감액 등 연간 약 5천만 원으로 예상 된다. (음식물 쓰레기 처리 절감 예상 비용 : 연간 쓰레기 예상 절감량 8만 kg * 200원 / kg = 16,000,000원, 식재료 구매 비용 예상 절감액 : 줄어드는 일 평균 예상 잔식량 약 40~50인분 * 3300원/1인 * 250일 = 33,000,000원 ~ 41,250,000원 등)

식수 인원 예측 알고리즘 개발을 통한 직접적인 효과 외에 추가적으로 발생하는 여러 간접적인 효과를 기대할 수 있다.

결론

제 1 절 시 사 점

본 연구는 S시청 단체 급식소의 실데이터를 기반으로 식수 예측 정확도 향상에 유의한 영향을 미칠 수 있는 다양한 변수를 탐색하고 그를 기반으로 다양한 기계 학습 기법을 적용한 식수 예측 모델을 제안하고 그 성능을 분석하였다.

그 결과 적합한 식재료의 주문, 이에 따른 사전적인 식재료 처리, 그리고 더 나아가 적합한 식수 예측 인원에 따른 식재료 주문, 조리 및 운영을 가능케 함으로써 미배식 잔반량을 줄이고 이에 따른 적합한 식수 예측의 비용 절감 및 절감된 비용을 활용하여 식사의 질을 개선하는데 사용하여 이용자의 만족도를 높이고 음식물 쓰레기로 인한 환경적 사회적 비용을 줄일 수 있는 선순환 구조의 기틀을 마련할 수 있었다.

본 연구는 학술적으로 다음 두 가지 측면에서 기존 연구들과 차별점이 있다고 본다.

첫째, 기존의 통일되지 않고 복잡한 한식 분류 체계에서 탈피하여 새로운 시각의 한식 메뉴 분류 방법을 제안하였다. 이를 통해 복잡한 한식 메뉴를 단순화할 수 있었고, 또한 본 연구에서 제안한 메뉴 분류가 신뢰성과 확장성을 토대로 향후 다른 집단 급식소에서의 식수 예측 모델링 적용이 가능하도록 하였다.

둘째, 실제 단체 급식소인 S시의 집단 급식소의 실제 데이터를 대상으로 분석하여 기계 학습을 도입, 데이터 과학에 기반한 식수 예측 모델링을 하여 성공한 최초의 사례 연구이다.

실무적으로는 다음과 같은 의의가 있다.

첫째, 즉, 다양한 기계 학습 방법을 적용하여 오차율을 현장에서 실제로 적용 가능한 수준으로 낮출 수 있었다. 이는 S 시청의 집단 급식소처럼 본 연구의 사례 연구 대상 외의 다른 집단 급식소에서도 적용할 수 있으며, 본 연구에서 제안된 한식 분류 체계에 따라 해당 급식소에 맞는 식수 인원 예측 모델링을 진행하면 해당 집단 급식소의 성공적인 식수 예측이 가능하다는 가능성을 보여준다.

둘째, 본 사례의 대상이 되는 기관은 제안된 모델을 활용하여 잔반을 줄일 수 있으며, 이를 통해 추가적인 비용 절감을 실현할 것으로 예상된다.

이는 유사한 타 기관의 급식 담당 의사결정자들에게도 시사점을 제공할 수 있다.

전체적으로 볼 때 음식물 쓰레기 처리 비용으로 매년 20조원이 사용되고 있으며, 이로 인해 발생하는 연간 온실 가스가 상당하다. 환경부에서 제공하는 음식물 쓰레기 발생현황에 따르면, 13년 기준으로 단체급식소가 배출하는 음식물쓰레기는 일 평균 171톤이며 다량 배출 사업장 중 집단 급식소가 차지하는 비율은 약 15%이다.

식수인원을 정확히 예측함으로써 잔식량을 줄여 단체급식소의 음식물쓰레기 양을 줄일 수 있고, 추후 다른 집단 급식소와 음식점에서 식수인원 예측 알고리즘 모델을 적용함으로써 음식물 쓰레기를 줄일 수 있을 것으로 기대한다.

셋째, 식수 예측에 따른 최적 생산으로 비용 최소화, 식수의 질 향상, 고객 만족도 증대 등으로 이어지는 선순환 효과를 가져 올 수 있게 하도록 음식물 쓰레기인 잔반을 줄이는 것 뿐 아니라 고객 만족을 위한 단체 급식소의 효율적 운영의 선순환 구조에 기여할 것을 기대한다.

넷째, 더 나아가 향후 고객 만족도 향상을 위한 새로운 영양 추천 메뉴 서비스의 기틀을 마련할 수가 있다. 이는 축적되는 데이터를 기반으로 제공되는 식사 메뉴의 품질을 지속적으로 높일 수 있기 때문이다

제 2 절 한 계 점

본 연구가 가지는 한계점은 다음과 같이 정리가 된다.

첫째, 현재 모델링의 한계점은 8개의 카테고리의 전체 설명 변수를 사용하였음에도 불구하고 이 변수들의 조합으로 다중회귀 모형을 적용하였을 때 아직 70% 이하의 설명력을 보인다는 한계가 있다. 여러 변수가 있음에도 극단적인 실제 값을 예측을 못하고 있어 주어진 데이터 만으로 알 수 없는 사항들이 존재한다고 할 수 있다.

주어진 정보 외에 식수 예측 모델링에 큰 변동을 주는 이벤트나 혹은 변수가 있을 것이며, 이를 찾아야 한다. 예측 가능한 범위를 벗어나는 식수 인원이 발생하고 있어, 추가적인 데이터 확보를 통해 이를 예측할 수 있는 추가적인 연구가 필요한 사항이다.

둘째, 예측 오차가 큰 Outlier에 해당하는 요일과 날짜에 대해 좀 더 세밀한 조사를 통해, Outlier에 해당하는 부분의 오차를 최소화 하는 연구가 필요하다고 할 수 있다.

Outlier 식별을 위한 Classifier를 만들어 학습을 통해 예상 보다 아주 많이 오거나 적게 오는 날을 Feature Engineering 기법을 바탕으로 새로운 변수로 추가 하여 Outlier에 대한 보정치를 다시 부여함으로써 오차율을 줄일 수 있는 추가 노력이 필요하다.

셋째, 메뉴에 대한 선호도 분석이 아직은 부족하다고 할 수 있어 이에 대한 보완 대책이 필요하다고 할 수 있다. 식수 가능 인원이 없는 경우에 대한 모델링 구현이 필요한 사항으로, 앞으로 본 연구에서 제시한 메뉴 선호 분석의 추가 개발 기법에 따라 예측 모델 고도화를 위한 추가 연구가 필요한 것으로 판단된다. 이는 데이터가 충분히 제공되지 못함으로써 미비했던 부분에 대해, 향후 더 많은 데이터를 가지고 식수 예측 모델링 관련 유효한 새로운 추가 설명 변수 내지 Feature Engineering에 의한 새로운 변수와 더 정교한 새로운 모델링 기법 개발의 필요성이 있다는 것을 보여 준다.

〈참고 문헌〉

강효성 (1994) : 환경 오염과 잔반 처리, 대한 영양사회 보수 교육 자료집, pp31~36

권수연, 이상목, 이영미 & 윤지현. (2010). 산업체 위탁급식소의 웰빙 메뉴 특성 및 현황
대한 영양사 협회학술지. 16(1): 1-12.

기상청 홈페이지 www.kma.go.kr

김세연 (2007) 중·고등학교 기술·가정 교과서에 게재된 음식의 종류와 분량 및 영양
적정성, 한국교원대학교 대학원 가정 교육학과, 석사 학위 논문

김종애. (2010). 위탁 급식업체의 급식 시설의 유형에 따른 서비스 만족도 비교, 단국
대학교 정보미디어 대학원 석사 학위 논문

김혜영. (2000). 산업체 급식소의 배식방법에 따른 음식 잔식량에 대한 실태 조사 연구.
한국 조리 과학회지. 16(5): 38-46.

나정기. (1994). 메뉴 계획과 디자인의 평가에 관한 연구

나정기. (2009). 메뉴 관리의 이해, 백산 출판사, 20, 서울.

박미자 (1996) “음식법(찬법)”의 조리학적 고찰, 대한 가정 학회지. Vol.34(2),
pp.283-302

박세정 (2017) 서울 지역 중학생의 환경인식과 학교급식 잔반실태 및 감량 교육에 관한
연구. 연세대학교 교육 대학원 : 영양 교육 전동 2017.2

박정소. (2011). 경기 지역 산업체 급식의 고객 만족도에 영향을 주는 요인에 관한 연구.
수원대학교 교육대학원 석사학위 논문

백옥희, 김미영 & 이복희. (2007). 산업체 근로자들의 급식메뉴에 대한 만족도 조사.
한국 식생활 문화 학회지. 22(4): 511-519.

송재성 (1996) 음식물 쓰레기 줄이기 정책, 단체 급식, 음식물 쓰레기 어떻게 줄일 것인가? 대한 영양사회 한국 영양학회 pp9~14p

심명교 김홍균 (1994) ; 음식물 쓰레기 감량화 규제에 대한 연구. 한국 환경 기술 개발원

양수일 (2016) 음식숙성과 와인숙성간의 상관관계, 세종대학교 관광 대학원 석사 학위
논문

양일선, 이보숙, 차진아 한경수 채인숙 이진미 (2011). 단체급식 (Food service in
institutions) 제3판. 파주 : 교문사, ix, 448 p

오정환. (1984). 호텔 캐터링 개론, 남영 문화사. p102

원용희. (1989). 현대 호텔 신당 경영론, 서울:대왕사

윤진숙 (1996) : 낭비적인 음식 문화 개선을 위한 교육 홍보 방안 단체 급식, 음식물
쓰레기 어떻게 줄인 것인가? 대한 영양사회 한국 영양학회 pp34~50

이민아 (2008) 한식 마케팅 모형 개발 연구: 한식 세계화 비전 및 전략안 [online].
<<http://kto.visitkorea.or.kr/file/download/bd/1329117638755.pdf,kto>>

이순목. 요인분석 I. 학지사, 1995.

이슬. (2010). 위탁급식 산업체 급식소의 점심식단 영양가 분석. 강릉원주대학교 교육대학원. 석사학위 논문.

이원갑, 김기진. (2014). 소비자의 인구 통계학적 특성에 따른 선호 메뉴 유형에 관한 연구 : 대구 · 경북을 중심으로, 한국조리학회지 Vol. 20, No. 1, pp89~104

이은용, 최순태 & 이수범. (2007). 산업체 급식 서비스 요인에 따른 이용자 만족도 분석. 호텔 관광연구. 26: 113-129.

이애주 (1996) 식음료 관리론, 일신사 pp179-181

이정임 (2010). 음식물류폐기물 민간처리시설의 환경오염 개선방안. 정책연구, 1-135.

이정임 김흥균 류주아 (1998) : 음식물 쓰레기 재활용 방안에 관한 연구. 경기 개발연구원, pp3-6

이정자 (1999) 메뉴 관리. 기문사. p8

이호현, 정승현, 최은정. "기계 학습 응용 및 학습 알고리즘 성능 개선 방안 사례 연구" 디지털 융복합 연구, 14.2, 2016, 246

이현주. (2013). 규제의 역설이 지배하는 급식 시장.. 대기업 막았더니 중견 기업이 독점. 한경 비즈니스, 통권 895호

이해영, 안선정, 양일선. (2004). 푸드코트형 산업체 급식소에서의 고객의 메뉴 선택 속성 규명. 대한 지역 사회 영양학회지. 9(2): 183-190.

임재영 (2015) 위탁급식 전문업체의 산업체 급식소 식수 예측 향상을 위한 식수 오차율 영향 요인 분석. 연세대학교 생활 환경 대학원 석사학위 논문

임혜진. (2008). 사업체급식소에 적합한 식수 예측기법 비교평가 및 선정. 연세 대학교 생활 환경 대학원 석사학위 논문.

장종근, 곽창근, 이부용, 오승용, 박성훈, 임병옥, 음식물 쓰레기로 버려지는 식량 자원의 경제적 가치 산정에 관한 연구, pp5 ~ 20, 한국 식품 개발 연구원, 2001

전무영 민혜선 (2000) 단체 급식소의 잔반량 감량을 위한 효율적인 방법에 관한 연구, 대한 지역 사회 영양 학회, 5(1), pp92~99

정라나. (2001). 대학교 급식의 운영 전략을 위한 식수 예측 모델 개발. 연세대학교 대학원 석사학위 논문.

정라나, 양일선, 백승희 (2003) 대학 급식소의 식수예측 기법 운영 현황, 한국 영양 학회지, November 2003, Vol.36(9), pp.966-973

정연경, 이미화 (2010) 음식 문화 NSDI 인터넷 자원 분류 체계 분석을 통한 한국 십진 분류법의 항목명 확장에 관한 연구 정보 관리 학회지, 27(4): 49-69

정연경, 최윤경 (2010) 음식 문화 분야의 DDC 분류 체계 개선 방안에 관한 연구 한국 비블리아 학회지 21(1), 2010.3, 43-57

정연경, 최윤경 (2011) 한식 분야의 듀이십진분류법 수정 전개 방안에 관한 연구. 한국 문헌 정보학회지 45(1), 2011.2, 29-49

정혜경 · 윤경수 · 김미혜. 2015. 「수운잡방」과 「음식디미방」에 나타난 조리법 비교. 한국 식생활 문화 학회지 30(1): 41-53, 2015

진양호, 1998. 메뉴 평가 모델의 개발에 관한 연구. 외식 경영 연구, p8~9

최순태. (2007) 산업체 위탁 급식 서비스 요인이 이용자 만족도에 미치는 영향에 관한 연구. 경희 대학교 관광 대학원 석사 학위 논문.

최현정 (2010) 메뉴 엔지니어링 및 판매량 추이 분석을 통한 메뉴 운영 전략 개발. 연세 대학교 생활 환경 대학원 석사 논문

홍성철 (1995) : 음식물 쓰레기 관리 정책 국민 영양 95-6:2-6

환경부 (1997) : 음식물 쓰레기 관리 정책 및 기술 동향과 감량화 자원화 실천 사례

환경부 (2009) : 『2008 전국 폐기물 발생 및 처리현황』

환경 정보 서비스 PUREUNURI (1996) : 음식물 쓰레기 퇴비화의 문제점과 개선 방안

한국 외식 연감 (2013)

한국 외식 정보 (2009). 「한국 외식 연감」, 한국 외식 정보

Adam, Jr. E.E. & Ebert, R.J. Production and operations management. Prentice-Hall Inc., New York, NY. 1978.

Adebanjo, D. Identifying problems in forecasting consumer demand in the fast moving consumer goods sector. *Benchmarking : An International Journal*. 7 (3):223- 230, 2000.

Ansel, D. & Dyer, C. A framework for restaurant information technology. *The Cornell Hotel, and Restaurant Administration Quarterly*. 40(3):74- 84, 1999.

Armstrong JS (1986): Research on forecasting: A quarter-century review, 1960-1984. *Interfaces* 16: 77-89

Chandler, S.J., Norton, L.C., Hoover, L.W. & Moore, A.N. Analysis of meal census patterns for forecasting menu item demand. *Journal of the America Dietetic Association*. 80:317- 323, 1982.

Cullen KO, Hoover LW, Moore, AN (1978): Menu item demand forecasting systems in hospital food service. *Journal of the America Dietetic Association* 73: 640-646

Davis M, Berger PD (1989): Sales forecasting in a retail service environment. *Journal of Business Forecasting* Winter, pp.8-17

DeFranco, A.L. The importance and use of financial forecasting and budgeting at the departmental level in the hotel industry as perceived by hotel controllers. *The Council on Hotel, Restaurant and Institutional Education*. 20(3):99- 110, 1997.

Dohrman, K. Critical is sues of the 1990's. *Journal of College & University*

Foodservice. 1(1):71- 74, 1993.

Dougherty D (1984): Forecasting Production Demand. In Rose, J.C. Handbook for health Care Foodservice Management. Aspen Systems Corp., Rockville, MD, pp.193-197

Edwards, J.S.A. & Ingram, H. Food, beverage and accommodation : An integrated operations approach. International Journal of Contemporary Hospitality Management, 7 (5):25- 28, 1995.

Farnum, N. R., & Stanton, L. W. (1989). *Quantitative forecasting methods*. Boston: PWS-Kent Publishing.

Gatsos, D. & Watkins, C. Transforming the campus dining experience. Food Management. 33 (4):36- 38, 1998.

Georgeoff, D.M. & Murdik, R.G. Manager' s Guide to Forecasting. Harvard Business Review. 64 (1):110, 1986.

Finley DH, Kim IY (1986): Use of selected management science techniques in health care foodservice systems. *Journal of Foodservice Systems* 4: 1-16

Georgeoff DM, Murdik RG (1986): Manager' s Guide to Forecasting. *Harvard Business Review* 64(1): 110

Haach, S.A. A study of patterning in meal participation of college students. NACUF Technical Bulletin. Spring, 1970.

Harris PJ (1995): A development strategy for the hospitality operations management curriculum. *International Journal of contemporary Hospitality Management* 7(5): 29-32

Hueter, J. & Swart, W. An Integrated Labor -Management System for Taco Bell. *Interface*. 28(1):75- 91, 1998.

Kisang Ryu, SooCheong (Shawn) Jang, Alfonso Sanchez (2003), Forecasting Methods and Seasonal Adjustment for a University Foodservice Operation

Levitt, E.D. The Psychology of Anxiety. Bobbs-Merrill Co., Inc., Indianapolis, IN. 1967.

Lin, B. S. & Vassar, J. A. (1992). Building a strategic forecasting system for hospital foodservice operations. *Journal of the American Dietetic Association*. 92(2): 204-207.

Liu CYD, Ridgway KA (1995): Computer-aided inventory management system-part 1: forecasting. *Integrated Manufacturing Systems* 6(1): 12-21

Mackle, M. & David, B.D. Developing a demand forecasting system for a foodservice operator. *Journal of the American Dietetic Association*. 68:456, 1976.

Magatet, M. & Beat rice, D.D. Developing a demand forecasting system for a foodservice operation. *Journal of American Dietetic Association*. 68: 456, 1976.

Mesersmith AM, Moore AN, Hoover LW (1978): A multi-echelon menu item forecasting system for hospitals. *Journal of the American Dietetic Association* 72: 509-515

Messersmith, A. M., & Miller, J. L. (1991). *Forecasting in Foodservice*. New York: John Wiley & Sons, Inc.

Messersmith AM, Miller JL (1992): *Forecasting in Foodservice*. John Willey & Sons, Inc., Canada

Miller JJ, McCahon CS, Miller JL (1991a): Forecasting production demand in school food service. *School Food Service Research Review* 15(2): 117-121

Miller JL (1990): Computer applications in foodservice management education in four-year hospitality management programs. *Hospitality Research Journal* 13(2): 16

Miller JL (1991): Forecasting in Foodservice: Surveys in three types of operations. *NACUFS Journal* 15

Miller JL, McCahon CS, Miller LL (1991b): Foodservice forecasting using simple mathematical models. *Hospitality Research Journal* 15(1): 43-58

Miller, J.L., McCahon, C.S. & Bloss, B.K. Food Production Forecasts with simple time series models. *Hospitality Research Journal*. 14(3) : 9- 22, 1991c.

Miller JL, McCahon CS, Miller LL (1993): Foodservice forecasting: differences in selection of simple mathematical models based on short-term and long-term data sets. *Hospitality Research Journal* 16(2): 95-102

Miller, J.L. & Shanklin, C.W. Forecasting menu - item demand in food service operations. *Journal of the America Dietetic Association*. 88:443- 449, 1988a.

Miller JL, Shanklin CW (1988): Forecasting menu-item demand in food service operations. *Journal of the American Dietetic Association* 88: 443-449

Pickert MJ, Miller JJ (1996): Food production forecasting in six commercial food service operation: A pilot study. *Hospitality Research Journal* 20(2): 137-144

Repko CJ, Miller JL (1990): Forecasting foodservice production. *Journal of the American Dietetic Association* 90: 1067

Robbins, S. P. (2000). *Managing Today* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.

Sanchez N, Miller JJ, Sanchez A, Brooks B (1995): Applying export systems technology to the implementation of a forecasting model in foodservice. *Hospitality Research Journal* 18(3): 25-38

SAS Institute, "Machine Learning: What it is & why it matters" , http://www.sas.com/en_us/insights/analytics/machine-learning.html (December 1, 2015)

Schmedgall, R.S . & DeFranco, A.L. Budgeting and forecasting. The Cornell Hotel, and Restaurant Administration Quarterly. 40(6):45- 51, 1999.

Scrimshaw, N.S . The Effect of stress on nutrition in adolescents and young adults. In F. P. Healded. Adolescent Nutrition and Growth. Appleton - Century Crofts, New York , NY. 1969.

Sill, B. & Decker, R. Applying capacity - management science. The Cornell

Hotel, and Restaurant Administration Quarterly. 40(3):22- 30, 1999.

Spears MC (2000): Foodservice Organizations: A managerial and systems approach, 4th ed., Prentice-Hall International Inc., Upper Saddle River, NJ

Spyros Makridakis, Steven C. Wheelwright and Rob J. Hyndman(1998) :
Forecasting: Methods and Applications, Third edition. John Wiley and Sons, 642pp,

Tanner, O. Stress. Time-Life Books, New York , NY. 1976.

Thompson, G. M. Labor scheduling, part 1. The Cornell Hotel, and Restaurant Administration Quarterly. 39 (5):22- 31, 1998a.

Thompson, G. M. Labor scheduling, part 2. The Cornell, Hotel, and Restaurant Administration Quarterly. 39 (6):26- 37, 1998b.

Thompson, G. M. Labor scheduling, part 3. The Cornell Hotel, and Restaurant Administration Quarterly. 40(1):86- 96, 1999a.

Thompson, G. M. Labor scheduling, part 4. The Cornell Hotel, and Restaurant Administration Quarterly. 40(3):85- 96, 1999b .

Waddell, D., & Sohal, A.S. Forecasting : The key to managerial decision making. Management Decision. 32(1):41- 49, 1994.

Wheelwright, S.C. & Makridakis, S. Forecasting Methods for Management. John Wiley & Son s , Inc., Canada. 1985.

Wood SD (1977): A model for statistical forecasting of menu item demand.
Journal of the American Dietetic Association 70: 254-259

Zifferblatt, A.M., Wilbur, C.S. & Pinsky, J.L. Influence on ecologic events on cafeteria food selections : Understanding food habits. *Journal of the American Dietetic Association*. 76: 9, 1980.



〈 감사의 글 〉

이

논문을

제가 존경하는

저의 어머니, 김필레 여사님과

사랑하는 아내, 성은에게 바칩니다

〈 Abstract 〉

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-

by Jongshik Jeon

Doctor of Philosophy in Management

Graduate School of Kyung Hee University

Advised by Dr. Ohbyung Kwon

Forecasting in food service is crucial because food is made just before eating and the food that remains is perishable. In order to validate the forecasting models for the numbers of the meals provided, selecting the relevant variables that affect these models is important.

This service was started in order to reduce the leftovers of the large cafeteria in Seoul City Hall. The main factors for leftovers were as follows: the leftovers before making the food, the leftovers after eating the food, and the non-distributed remaining food. Among these factors, the most influential factor was non-distributed remaining food. Leftovers before making the food and the leftovers after eating the food are controllable with the support of the employees of Seoul Metropolitan City, which would result in the total leftovers of Seoul City cafeteria being relatively limited.

The non-distributed remaining food, however, was what caused a large gap between the number of expected persons and the number of actual meals served, with the number of actual meals served being much lower than expected.

The variables for forecasting the numbers of the actual meals served consisted of two kinds of factors: internal factors such as the daily menu, the number of available persons (number of persons in attendance, the number of persons on a business trip, and the number of persons off-duty), the daily number of actual meals served, the characteristics of the employees, officers, staff, etc.; and external factors such as weather, temperature, holidays, seasonal factors, etc.

The data for this study consisted of 3 years of available actual data on the Seoul Metropolitan City cafeteria (the details of use) from May 01, 2015 to April 30, 2018. This data included the daily number of the actual persons provided with meals, the daily amount of leftovers, the daily menu details, the daily available number of persons for the meals (the number of people in attendance, the number of person on a business trip, and the number of persons off-duty), the daily weather and temperature conditions, etc.

Data manipulation for the daily menu is important, and the assortment and grouping method of the menu follows the Dewey Decimal Classification for Korean foods through which hundreds of dishes are classified, grouped, and programmed.

The applicable modeling method is multi regression analysis and machine learning method, from which we will select the best forecasting results.

The objective of this study is to make a forecasting model for the number of meals served at the Seoul Metropolitan Cafeteria in order to reduce leftovers.

keywords : R, machine learning, leftovers, large dining places, forecasting models for number of persons for meals, overproduct, plate waste

