

기계학습 알고리즘 기반 재활 센터 환자 노쇼 예측 모델 개발

전성은¹, 박지민², 신준용³, 서혜린⁴, 조석현*
한남대학교¹, 국립금오공과대학교², 건국대학교³, 경기대학교⁴
*University of California, San Diego (UCSD)

castle1122eun@gmail.com, parkjimin000409@gmail.com, wnsdyd0504@gmail.com
shr200106@gmail.com, *justinshcho@gmail.com

Development for Machine Learning-powered Patient No-show Prediction Model at a Rehabilitation Center

Seongeun Jeon¹, Jimin Park², Junyong Shin³, Hyerin Seo⁴, Seokheon Cho*
¹Hannam Univ., ²Kumoh National of Technology, ³Konkuk Univ., ⁴Kyonggi Univ.
*University of California, San Diego (UCSD)

요약

Patient no-shows significantly affect healthcare efficiency and resource allocation in rehabilitation centers. This paper presents a machine learning approach to predict if a patient will show up for their scheduled appointment using a Hospital Appointment Management (HAM) dataset. This dataset is generated by integrating patient records from Mendeley Data at a rehabilitation center with weather data collected from Brazil's National Institute of Meteorology. A two-stage under-sampling method combining Tomek Link and Random Under-Sampling was applied to the training dataset to deal with a highly imbalanced dataset. Three machine learning algorithms, such as Logistic Regression, K-Nearest Neighbor, and Random Forest, were utilized to provide the patient no-show prediction models. The Random Forest-based patient no-show prediction model outperformed other models based on the performance results, especially in terms of Recall, which is the proportion of actual patient no-shows that were correctly classified as the case when a patient fails to show up for a scheduled appointment.

I. 서론

의학 기술의 발전과 의료 보장성 강화 정책의 시행으로 현대인의 평균 수명이 증가하고, 이에 따라 환자 수 또한 늘어나고 있다. 특히, 상급종합병원의 진료 절차가 간소화되면서 환자들은 우수한 의료진과 시설을 갖춘 병원을 선호하게 되었으며, 이는 외래진료 예약의 필요성을 더욱 부각시키고 있다 [1]. 대부분의 의료기관에서는 외래진료의 70~90%가 예약제로 운영되며, 이를 통해 환자 와 의료기관 모두가 효율적인 진료 제공의 이점을 누릴 수 있는 것이다 [2, 3]. 그러나, 환자가 진료 예약을 하였지만 진료 시간에 오지 않는 노쇼 (no-show) 문제는 많은 부작용을 낳고 있다. 노쇼는 병원 경영의 효율성을 감소시키고 다른 환자들에게 있어서는 진료 기회 상실과 이에 장기적으로는 고객의 이탈을 유도할 수도 있다. 또한, 노쇼 환자들을 고려하여 많은 환자들을 초과예약제를 시행함에 따라 환자들의 대기시간 증가로 인해 고객 만족도가 저하될 수도 있다. 이 뿐만 아니라 노쇼 환자들까지 고려한 병원 예약 시스템을 운영하기 위해 병원에서는 불필요한 인력 낭비를 초래하며 불필요한 비용 손실도 야기하게 된다 [4-7]. 한국현대경제연구원에 따르면, 병원에서 노쇼로 인한 매출 손실이 연간 약 2조 4,980억 원에 달해 서비스 업종 중 가장 큰 손실을 기록하고 있다고 한다 [8]. 이에 본 연구는 환자 정보, 환자 예약 정보와 날씨 데이터를 활용하여 환자의 노쇼 발생 여부를

예측하고 해결 방안을 모색하고자 한다.

노쇼를 예측하는 데 있어 날씨 데이터를 활용한 연구가 활발히 이루어지고 있다. D. Liu *et al.* 은 미국 보스턴 소아병원에서 161,822건의 환자 정보와 날씨 데이터를 수집하여 Logistic Regression (LR)과 Artificial Neural Network (ANN)알고리즘을 적용하였으며, ANN 가 더 높은 예측 성능을 보였다 [9]. 또 다른 연구에서는 브라질 대학 병원에서 날씨 데이터를 바탕으로 Random Forest (RF) 기반 환자 노쇼 발생 여부 모델이 가장 우수한 성능을 가져옴을 보였다 [10]. 진료 예약을 한 환자들 대부분은 예약을 한 시간에 진료를 받기 때문에, 노쇼 관련 데이터는 불균형 데이터라 할 수 있다. 기존 연구들은 over-sampling을 통해 데이터 불균형 문제를 해소하고자 하였다. 하지만, 이 방식은 인위적으로 생성된 데이터가 학습에 영향을 미치기에 성능의 신뢰성을 저하시킬 수 있기 때문에 실제 병원 예약 시스템에 적용 시 한계가 있다. 따라서, 본 연구는 이러한 한계를 보완하고자 두 단계 언더 샘플링 (two-stage under-sampling) 기법을 적용한 데이터셋에 대해서 모델 학습을 진행하고 샘플링 기법을 적용하지 않은 실제 데이터 세트로 검증하여 보다 신뢰성 있는 예측 모델을 구축하고자 한다.

본 논문은 다음과 같이 구성된다. 제 II 장에서는 연구에서 사용된 데이터셋과 전처리 과정을 설명한다. 제 III 장에서는 환자 노쇼 예측 모델을 위해 사용한 기계학

습 알고리즘들과 예측 모델들의 성능을 분석하기 위한 성능 지표를 소개한다. 제 IV 장에서는 기계학습 기반 환자 노쇼 예측 모델 별 성능에 대한 분석을 수행하였다. 마지막으로 제 V 장에서는 본 연구의 결론과 향후 연구 과제를 제시하며 마무리한다.

II. 데이터 세트 및 전처리 과정

본 논문에서는 Mendeley Data에서 제공하는 환자 데이터세트와 브라질 국립 기상 연구소 INMET (Instituto Nacional de Meteorologia)에서 제공하는 날씨 데이터세트를 결합하고 가공하여 활용하였다 [11, 12]. 데이터 세트의 종속변수는 환자의 노쇼 (No-show) 여부이다. 환자 데이터세트의 경우, COVID-19 팬데믹의 영향을 배제하기 위해 2019년 11월 30일 이후의 데이터는 포함하지 않았다. 또한, 환자 데이터세트와 날씨 데이터세트간 중복된 정보를 지닌 변수는 하나로 병합하였으며, 대다수가 결측 값을 가진 변수들은 성능 저하를 야기할 수 있기 때문에 삭제하였다.

표 1은 전처리 후 생성한 병원 예약 관리 (Hospital Appointment Management: HAM) 데이터세트로 82개의 독립변수들과 1개의 종속변수를 보여주고 있다. HAM 데이터세트의 데이터 샘플의 총 개수는 31,030개이다.

독립변수들 중에서 다음과 같은 날씨 관련 특성 (feature)들이 포함되어 있다: 온도 (temp), 이슬점 온도 (dew_temp), 강수량 (precipitation), 압력 (pressure), 일사량 (solar), 습도 (humidity), 풍속 (wind_speed) 그리고 최대 풍속 (max_wind). 이 날씨 관련 특성들에 대해서는 해당 진료 예약 시간대를 기준으로, 시간대의 1시간, 3시간 및 5시간 이전에 측정된 데이터를 독립변수로 추가하였다. 또한, 해당 진료 예약 시간대에서 과거 3시간 동안 그리고 5시간 동안의 평균값을 독립변수로 추가하였다. 마지막으로 해당 진료 예약 일자의 전날의 평균값과 최대값 변수를 추가적으로 생성하였다. 예를 들어, 온도의 경우 진료 예약 시간대의 1시간, 3시간 및 5시간 이전 온도를 각각 temp_at_1hr, temp_at_3hr 그리고 temp_at_5hr라 정의하였다. 진료 예약 시간대에서 과거 3시간 동안의 평균값과 5시간 동안의 평균값은 각각 temp_avg_3hr와 temp_avg_5hr라고 명명하였다. 또한, 진료 예약 일자의 전날 평균 온도와 최대 온도를 각각 temp_avg_1d와 temp_max_1d라 정의하였다.

독립변수들 중에서 다음과 같은 환자 예약 관련 특성들이 존재한다: 성별 (gender), 정신적 장애 (intellectual) 또는 행동 장애 (motor)로 구분되는 장애 종류 (disability), 환자 간병인 필요 여부 (assist), 환자 모니터링 필요 여부 (enf), 작업치료과 진료 여부 (occupational therapy), 교육치료과 진료 여부 (pedagogical therapy), 물리치료과 진료 여부 (physiotherapy), 심리치료과 진료 여부 (psychotherapy), 특정 진료과목이 없는 일반 진료 여부 (sem especialidade), 소통 장애에 대한 치료 여부 (speech therapy), 보호자 동행 필요 여부 (needs_companion), 나이 (age), 환자가 병원에 첫 방문한 날짜로부터 다음 진료 예약 날짜까지의 소요된 일자 (after_first_visit), 진료 예약 년도 (app_year), 진료 예약 달 (app_month), 해당 병원이 평일 진료만을 하기 때문에 월요일을 0으로 설정하여 평일을 숫자로 변환한 변수인 진료 예약 요일 (app_dayweek), 병원 첫 진료 시작 시간인 7시를 기준으로 진료 예약 시간이 몇 분이나 지났는지를 나타내는 변수인 진료 시간 오프셋 (offset_minute), 그리고 진료 예약 시간을 오전과 오후로 구분한 변수인 진료 예약 시간대 (app_shift)들로 이루어져 있다.

표 1. HAM 데이터세트의 데이터별 타입 및 단위

Dependent Variable		
Variable	Type	[Unit] / Value
appointment_status	Integer	{1 (no-show), 0 (show-up)}
Independent Variables		
Variables	Type	[Unit] / Value
temp	Double	[°C]
dew_temp	Double	[°C]
precipitation	Double	[mm]
pressure	Double	[mB]
solar	Double	[KJ/m2]
humidity	Double	[%]
wind_speed	Double	[m/s]
max_wind	Double	[m/s]
gender	Integer	{1 (Female), 0 (Male)}
disability	Integer	{1 (Intellectual), 0 (Motor)}
assist	Integer	{1 (Yes), 0 (No)}
enf	Integer	{1 (Yes), 0 (No)}
occupational therapy	Integer	{1 (Yes), 0 (No)}
pedagogical therapy	Integer	{1 (Yes), 0 (No)}
physiotherapy	Integer	{1 (Yes), 0 (No)}
psychotherapy	Integer	{1 (Yes), 0 (No)}
sem especialidade	Integer	{1 (Yes), 0 (No)}
speech therapy	Integer	{1 (Yes), 0 (No)}
needs_companion	Integer	{1 (Yes), 0 (No)}
age	Integer	-
after_first_visit	Integer	-
app_year	Integer	{2016, ..., 2019}
app_month	Integer	{1, ..., 12}
app_dayweek	Integer	{0, ..., 4}
offset_minute	Integer	{0, ..., 680}
app_shift	Integer	{1 (Morning), 0 (Afternoon)}

그림 1은 본 연구에서 고려하는 HAM 데이터세트의 종속 변수인 appointment_status의 클래스별 구성을 보여주고 있다. 다수 클래스인 show-up (= 0)과 소수 클래스인 no-show (= 1)에 해당하는 데이터 샘플 수는 각각 28,634개와 2,396개이다. 이는 다수 클래스와 소수 클래스의 비율이 92.28%와 7.72%로 극심한 불균형 데이터임을 보여주고 있다. 데이터 불균형 문제를 해결하여 환자 노쇼 예측 모델의 성능을 향상시키기 위해서는 적절한 샘플링 기법이 필요함을 알 수 있다.

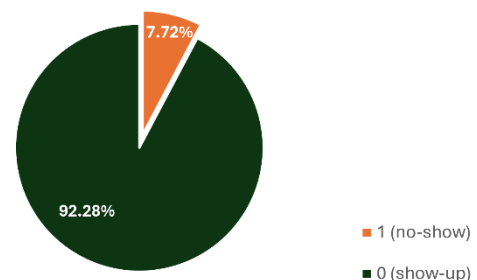


그림 1. HAM 데이터세트의 클래스 구성

III. 알고리즘 및 성능 평가 지표

본 연구에서는 환자 노쇼 예측 모델을 제시하기 위해 Logistic Regression (LR), K-Nearest Neighbor (KNN) 및 Random Forest (RF) 등 3가지 기계학습 알고리즘을 사용하였다. LR은 이진 분류 문제에서 특정 사건이 발생할 확률을 예측하는 알고리즘이다. 종속 변수가 이산적이며 Sigmoid 함수로 결과값을 0에서 1 사이로 제한하여 특정 클래스에 속할 확률을 계산한다. 본 연구에서 사용된 LR 알고리즘의 하이퍼파라미터는 허용 가능한 오차 범위인 엡실론 (epsilon)은 1.0E-6이고 다음 단계로의 스텝 사이즈 (step size)는 0.001로 설정하였다. KNN은 새로운 데이터 포인트의 클래스를 주변 K개의 데이터 포인트들의 다수결로 결정하여 분류를 수행하는 알고리즘이다. KNN의 하이퍼파라미터로 이웃의 수 (K)를 1로 설정하였다. RF는 여러 결정 트리를 결합하여 예측을 수행하는 앙상블 학습 알고리즘이다. 이 RF 기반 모델의 하이퍼파라미터로 최대 트리 깊이 (tree depth)는 32이고 잎 노드 (leaf node)의 최소 데이터 샘플 수 (minimum node size)는 6으로 설정하였다.

환자 노쇼 예측 모델의 성능 평가 시 재현율 (Recall), 정확도 (Accuracy) 및 F1-score를 주요 지표로 사용하였다. 특히 의료 환경에서 노쇼 예측의 특성상 실제 노쇼 환자를 틀리지 않고 예측하는 것이 중요하므로, 실제 노쇼 환자를 정확히 예측하는 모델 성능을 측정하는 Recall을 가장 중요한 지표로 고려하였다. 각 평가 지표 값은 0과 1 사이의 값을 가지며 1에 가까울수록 모델의 성능이 우수하다고 할 수 있다.

IV. 환자 노쇼 예측 모델 성능 분석

4.1 환자 노쇼 예측 모델 개발 과정

그림 2는 본 연구에서 개발한 환자 노쇼 예측 모델의 전체 개발 흐름도를 도시하고 있다. 먼저, HAM 데이터 세트를 학습 및 검증 (Training and Validation) 데이터 세트와 최종 테스트 (Test) 데이터 세트로 각각 80%와 20%로 분할하였다. 이 80%의 데이터 세트를 다시 80%와 20%로 분할하여 학습 데이터 세트와 검증 데이터 세트로 사용하였다. 이로써 학습 데이터 세트와 검증 데이터 세트는 전체 HAM 데이터 세트의 각각 64%와 16%를 차지한다. 이 학습 데이터 세트와 검증 데이터 세트에 대해 K=5인 K-겹 교차 검증 기법 (K-fold cross validation)을 수행하여 본 연구에서 고려하는 3개의 기계학습 기반 예측 모델들에 대해 최적의 하이퍼파라미터를 찾았다. 특히, 환자 노쇼 예측 모델의 성능을 향상시키기 위해서 학습 데이터 세트에 대해 데이터 불균형을 해소하는 샘플링 기법을 적용하였다. 최종적으로, 최적의 하이퍼파라미터가 설정된 환자 노쇼 예측 모델을 이용해 테스트 데이터 세트의 성능을 평가하였다.

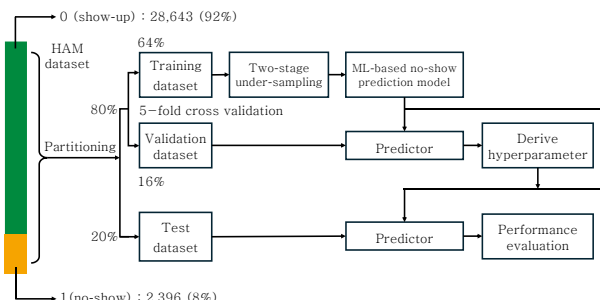


그림 2. 환자 노쇼 예측 모델 전체 흐름도

4.2 불균형 데이터 처리를 위한 두 단계 언더 샘플링 기법

그림 1에서 확인했던 것처럼, 본 연구에서 고려하는 HAM 데이터 세트는 극심한 불균형을 특징으로 하고 있다. 데이터 불균형은 개발하고자 하는 예측 모델의 성능을 감소시키는 악영향을 끼친다. 데이터 불균형 문제를 해결하기 위해 학습 데이터 세트에 대해서만 Tomek Link (T-link)와 Random Under-Sampling (RUS)을 결합한 두 단계 언더 샘플링 (two-stage under-sampling) 기법을 적용하였다. T-Link는 서로 다른 클래스에 속하는 두 샘플 간의 거리가 최소인 쌍을 찾아, 그 중 다수 클래스에 속하는 샘플을 제거함으로써 클래스 간 경계를 더욱 명확히 만드는 기법이다. 학습 데이터 세트에 T-Link 기법을 우선 적용하여 클래스 간의 데이터 경계를 정제하고, 데이터 불균형 문제를 개선함으로써 학습 모델이 클래스 간 차이를 더 명확히 학습할 수 있도록 하였다. 이 후 RUS 기법을 사용하여 다수 클래스의 데이터 샘플들을 무작위로 제거하여 소수 클래스의 샘플 수와 동일하게 균형을 맞추었다. 즉, 두 단계 언더 샘플링을 통해 학습 데이터 세트에 한해서 원래 다수 클래스인 0 (show-up)의 데이터 샘플 수와 소수 클래스인 1 (no-show)의 데이터 샘플 수를 1:1 비율로 균형을 맞춘 것이다. RUS 기법은 무작위로 데이터를 제거하기 때문에 중요한 정보가 손실될 위험이 있으며, 이로 인해 성능 저하가 발생할 수 있다. 따라서 두 단계 언더 샘플링 기법은 RUS를 단독으로 사용하는 방식 대신 T-Link를 우선 적용함으로써, 다수 클래스의 불필요한 정보 손실을 최소화하고 클래스 불균형 문제를 효과적으로 해결할 수 있는 장점을 가진다.

4.3 환자 노쇼 예측 모델 성능 평가

그림 3은 3개의 기계학습 기반 환자 노쇼 예측 모델 성능 결과를 시각화한 것이다. 본 연구에서 중점적으로 다루는 지표는 실제 노쇼 환자에 대해서 노쇼를 할 것이라고 올바르게 예측하는 지표인 Recall이다. 해당 성능 지표에서 RF 기반 모델은 0.729로 가장 높은 성능을 보였다. RF 기반 예측 모델은 Recall뿐만 아니라 Accuracy에서 또한 다른 기계학습 모델들보다 우수하였다. 그러나 RF 기반 환자 노쇼 예측 모델에서 Recall이 0.729인 것에 비해 Accuracy가 0.612로 낮은 이유는, 모델이 소수 클래스인 실제 노쇼 환자에 대해서는 높은 예측 정확도를 보이지만, 다수 클래스인 정상 도착 환자의 예측 정확도가 낮기 때문이다. 즉, 노쇼 환자에 대한 예측 성능은 좋지만, 정상 도착 환자에 대해서는 오류가 더 많음을 나타낸다. 또한, F1-score도 RF 기반 모델이 가장 좋지만 0.225로 상당히 낮은 값이다. 여기서 F1-score 값이 낮은 이유는 환자 노쇼 예측 모델의 예측 성능이 열악하다기 보다는 본 연구에서 고려하는 HAM 데이터 세트의 극심한 데이터 불균형에 기인한다.

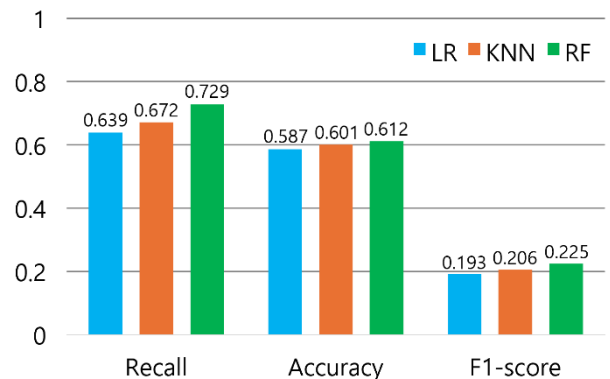


그림 3. 환자 노쇼 예측 모델 성능 결과

표 2는 테스트 데이터셋에 대해 RF 기반 환자 노쇼 예측 모델로 예측한 결과를 나타낸 혼동 행렬(confusion matrix)이다. 그림 3에서 긍정 클래스(positive class: no-show)인 소수 클래스의 Recall 값이 0.729임에도 불구하고 F1-score 값이 0.225로 낮은 이유는 정밀도(Precision)가 낮기 때문이다. 표 2에서 확인할 수 있듯이, Precision 값은 0.133으로 낮게 나타났지만, 이는 예측 모델 자체의 성능이 나쁜 것을 의미하지는 않는다. 이는 특이도(Specificity)와 Recall 각각 0.602와 0.729로 그렇게 작은 값이 아니기 때문이다. 다시 말해, 본 연구에서 개발한 RF 기반 예측 모델은 실제 show-up인 다수 클래스에 해당하는 환자들에 대해서는 60.2%의 정확도로 예측을 수행하며, 실제 no-show인 소수 클래스 환자들에 대해서는 72.9%의 정확도로 예측을 수행한다. 하지만 그림 1에서 보여준 것처럼, 극심한 데이터 불균형으로 인해 no-show로 예측된 환자 중 실제로 no-show한 환자는 349명에 불과한 반면, 실제로 show-up한 환자는 2,279명에 이르기 때문에 Precision 값이 0.133으로 매우 낮게 나타났다.

따라서, HAM 데이터셋과 같이 극심한 데이터 불균형을 가진 환경에서는 F1-score 값보다는 실제 no-show 환자를 올바르게 예측하는 비율인 Recall에 더 큰 가중치를 두는 것이 중요하다. 특히, RF 기반 환자 노쇼 예측 모델은 약 72.9%의 정확도로 실제 no-show 환자를 가장 잘 예측하는 우수한 모델이다.

표 2. RF 기반 환자 노쇼 예측 모델 혼동 행렬

		Predicted	
		1 (no-show)	0 (show-up)
Actual	1 (no-show)	349 (TP)	130 (FN)
	0 (show-up)	2,279 (FP)	3,450 (TN)

V. 결론

본 연구에서는 병원 노쇼 예측을 위해 Mendeley Data에서 제공하는 환자 데이터셋과 브라질 국립 기상 연구소(INMET)에서 제공된 기상 데이터셋들로부터 생성한 병원 예약 관리(Hospital Appointment Management: HAM) 데이터셋을 사용하여 환자 노쇼 예측 모델을 개발하였다. Tomek link와 Random Under-Sampling을 결합한 두 단계 언더 샘플링(two-stage under-sampling) 기법을 적용하여 학습 데이터셋에 대한 불균형 문제를 해결하였다. 테스트(Test) 데이터셋의 경우 샘플링 기법을 적용하지 않은 실제 데이터로 개발한 환자 노쇼 예측 모델들에 대한 성능 평가를 진행하였다. Logistic Regression(LR), K-Nearest Neighbor(KNN) 및 Random Forest(RF) 등 3개의 기계학습 알고리즘 기반 예측 모델을 개발하였다. 이 중 Random Forest 기반 환자 노쇼 예측 모델이 Recall 값과 Accuracy 값이 각각 0.729와 0.612로 가장 우수한 성능을 보였다. 따라서, 두 단계 언더 샘플링 기법을 학습 데이터셋에만 적용한 후 RF 기반 환자 노쇼 예측 모델로 학습한 경우가 가장 우수한 성능을 보인 것이다.

향후 연구에서는 비선형 데이터 패턴을 효과적으로 학습할 수 있는 인공지능 알고리즘을 추가적으로 도입, HAM 데이터셋으로부터 주요 특성들을 선택(feature selection) 또는 다양한 샘플링 기법 적용 등을 통하여 환자 노쇼 예측 모델에 대한 성능을 향상시키고자 한다.

ACKNOWLEDGMENT

Following are results of a study on the "Leaders in Industry-university Cooperation 3.0" Project, supported by the Ministry of Education and National Research Foundation of Korea.

This work was supported by "University Innovation Grant" from the Ministry of Education and National Research Foundation of Korea.

This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the National Program for Excellence in SW supervised by the IITP (Institute of Information & communications Technology Planning & Evaluation) (2021-0-01393).

참고 문헌

- [1] Y. Shin, H. Kang, N. Kim, Y. Jeong, D. Kim, and D. Hwang, "Improving the Sustainability of Healthcare Delivery in Line with NHI Coverage Enhancement," *Korea Institute for Health and Social Affairs*, vol. 8, no. 142, pp. 1-163, Jul. 2013.
- [2] D. Shin, "Understanding and Forecasting Hospital No-shows," *Korea University Graduate School of Public Health, Master's Thesis*, pp. 1-45, Jun. 2005.
- [3] C. DuMontier, K. Rindfleisch, J. Pruszynski, and J. J. Frey III, "A Multi-method Intervention to Reduce No-shows in an Urban Residency Clinic," *Journal of the American Board of Family Medicine*, vol. 45, no. 9, pp. 634-641, Oct. 2013.
- [4] S. Yoon, "An Exploratory Study for Preventing No-show of Tourism Companies," *International Journal of Tourism and Hospitality Research*, vol. 28, no. 1, pp. 21-35, Jan. 2014.
- [5] J. Shim, Y. Lee, S. Lee, S. Park, and T. Kim, "An Analysis of Factors Related to Delayed Arrival of Appointment Time among Outpatients of the Department of Psychiatry," *Journal of Health Informatics and Statistics*, vol. 40, no. 3, pp. 85-99, Sep. 2015.
- [6] J. Lim, S. Lee, T. Kim, and J. Kim, "Reducing Appointment No-shows in Hospitals," *Korea Journal of Hospital Management*, vol. 22, no. 4, pp. 50-60, Oct. 2017.
- [7] M. J. Hashim, P. Franks, and K. Fiscella, "Effectiveness of Telephone Reminders in Improving Rate of Appointments Kept at an Outpatient Clinic: A Randomized Controlled Trial," *Journal of the American Board of Family Practice*, vol. 14, no. 3, pp. 193-196, May. 2001.
- [8] Korea Economic Research Institute, "Current Status of No-shows," 2016.
- [9] D. Liu, W. Shin, E. Sprecher, K. Conroy, O. Santiago, G. Wachtel, and M. Santillana, "Machine Learning Approaches to Predicting No-shows in Pediatric Medical Appointment," *npj Digital Medicine*, vol. 5, no. 1, pp. 1-11, Apr. 2022.
- [10] L. H. A. Salazar, V. R. Q. Leithardt, W. D. Parreira, A. M. da Rocha Fernandes, J. L. V. Barbosa, and S. D. Correia, "Application of Machine Learning Techniques to Predict a Patient's No-show in the Healthcare Sector," *Future Internet*, vol. 14, no. 1, pp. 1-21, Dec. 2021.
- [11] Mendeley Data, "Medical Appointments No-show", 2023, (<https://doi.org/10.17632/wm6w2fvkfy.1>)
- [12] Instituto Nacional de Meteorologia (INMET), Brazilian National Institute of Meteorology Data, Jul. 2024, (<https://portal.inmet.gov.br/>).