

# 기계학습 기반 지하매설물 속성 및 밀집도를 활용한 지반함몰 위험도 예측 모델

## Ground Subsidence Risk Grade Prediction Model Based on Machine Learning According to the Underground Facility Properties and Density

이 성 열<sup>1)</sup> · 강 재 모<sup>†</sup> · 김 진 영<sup>2)</sup>

Sungyeol Lee · Jaemo Kang · Jinyoung Kim

Received: February 15<sup>th</sup>, 2023; Revised: February 23<sup>rd</sup>, 2023; Accepted: March 9<sup>th</sup>, 2023

**ABSTRACT** : Ground subsidence shows a mechanism in which the upper ground collapses due to the formation of a cavity due to the movement of soil particles in the ground due to the formation of a waterway because of damage to the water supply/sewer pipes. As a result, cavity is created in the ground and the upper ground is collapsing. Therefore, ground subsidence frequently occurs mainly in downtown areas where a large amount of underground facilities are buried. Accordingly, research to predict the risk of ground subsidence is continuously being conducted. This study tried to present a ground subsidence risk prediction model for two districts of ○○ city. After constructing a data set and performing preprocessing, using the property data of underground facilities in the target area (year of service, pipe diameter), density of underground facilities, and ground subsidence history data. By applying the dataset to the machine learning model, it is evaluated the reliability of the selected model and the importance of the influencing factors used in predicting the ground subsidence risk derived from the model is presented.

**Keywords** : Ground subsidence, Underground facilities, Density of underground facilities, Machine learning, Ground subsidence risk prediction model

**요 지** : 지반함몰의 주요 발생원인은 지하매설물의 손상으로 알려져 있다. 지반함몰은 상·하수관의 손상으로 인한 물길 형성에 따른 지반 내 토립자의 이동으로 공동이 형성되어 상부지반이 붕괴되는 메커니즘을 보이고 있다. 따라서 지반함몰은 지하매설물의 밀집도가 높은 도심지를 중심으로 발생하고 있으며, 사고 발생 시 인명 및 경제적 피해를 야기하므로 사고에 대한 대비가 반드시 필요하다. 이에 따라 지반함몰 위험을 예측하기 위한 연구가 꾸준히 수행되고 있으며, 본 연구에서는 ○○시의 2개 구를 대상으로 지반함몰 위험도 예측 모델을 제시하고자 하였다. 대상 지역의 지하매설물 속성 데이터(활용년수, 관직경)와 지하매설물 밀집도, 지반함몰 이력 데이터를 활용하여 데이터셋을 구축하고 전처리를 수행한 뒤, 기계학습 모델에 적용하여 최적의 평가지표가 도출되는 모델을 선정하였으며, 선정된 모델의 신뢰도를 평가하고 모델에서 도출되는 지반함몰 위험도 예측 시 활용된 영향인자의 중요도를 제시하고자 하였다.

**주요어** : 지반함몰, 지하매설물, 밀집도, 기계학습, 지반함몰 위험 예측모델

## 1. 서 론

지반함몰의 주요 원인은 지하매설물 손상으로 인한 물길의 형성에 따른 지반 내 토립자의 이동 시 생성되는 공동의 확장으로 알려져 있다(Kim et al., 2017). 지하매설물은 인구의 밀집도가 높은 도심지에 집중적으로 매설되어 있기 때문에 도심지에서의 지반함몰 발생 빈도가 높아 사고 발생 시 커다란 사회적 혼란을 야기할 수 있다(Lee et al., 2022). 따라서 지반함몰의 근본적인 원인 및 메커니즘 분석 등을 통해 사고를 예방해야 하며, 이와 관련된 연구도 꾸준히 진행되고 있다.

지반함몰 발생의 전조현상인 지반 내 공동의 발생 메커니즘 규명을 위해 Kuwano et al.(2006)은 일본의 표준사를 활용하여 실내 모형실험을 실시하였으며, Mokunoki et al.(2009)은 모형토조 하부에 하수관 균열을 모사하여 지반 내 공동 발생 메커니즘을 확인하고 X-ray와 CT 등의 장비를 통한 공동 발생의 시각화에 대한 연구를 발표하였다. 서울시(2014)에서는 지반함몰 발생 현황을 조사하여 사고 발생의 주요 원인을 상수관 및 하수관의 손상으로 발표하였으며, Takeuchi et al.(2017)은 기계학습 알고리즘인 Decision Tree를 활용하여 지반함몰에 영향을 미치는 영향 인자에 대해 선정하는

1) Department of Geotechnical Engineering Research, Korea Institute of Civil Engineering and Building Technology, Postdoctoral Researcher

† Department of Geotechnical Engineering Research, Korea Institute of Civil Engineering and Building Technology, Senior Researcher (Corresponding Author : [jmkang@kict.re.kr](mailto:jmkang@kict.re.kr))

2) Department of Geotechnical Engineering Research, Korea Institute of Civil Engineering and Building Technology, Senior Researcher

연구를 발표하였다. 또한, Jin(2018)은 AHP 분석을 통해 지반함몰 영향인자들의 가중치 분석을 실시하였다.

지반함몰 발생 위험을 예측하는 연구도 꾸준히 수행되고 있다. Han(2017)은 지반함몰의 주요 원인인 하수관로를 중심으로 조사된 CCTV 자료와 지중탐사레이더의 공동 탐사 자료를 활용하여 지반함몰 위험도 평가에 관한 연구를 발표하였고, Kim(2018)은 로지스틱 회귀분석을 통한 지반함몰 위험도 회귀식을 제시하였다. 또한, Lee et al.(2022)는 지하매설물의 속성 값 중 활용년수, 관로 직경을 영향인자로 선정하여 기계학습을 통한 지반함몰 위험도 예측 모델을 선정하고 위험지도를 제시하였다.

이와 같이, 지반함몰 사고를 방지하기 위해 다양한 접근 방법의 위험도 예측 연구가 수행되고 있으나, 광범위한 지역을 대상으로 다양한 요인에 의해 복합적으로 발생하는 지반함몰 특성 상 높은 수준의 정확도 및 신뢰도를 나타내는 결과를 도출하기는 어려운 실정이다.

따라서 본 연구에서는 ○○시 ○○구를 대상으로 지하매설물 속성정보 중 활용년수, 관경, 관로 길이와 지반함몰과의 높은 상관성이 예상되는 관로의 밀집도를 영향인자로 선정하여 지반함몰 위험도에 대한 기계학습 예측모델을 제시하고자 하였다. 이를 위해 다양한 조건의 데이터셋을 적용한 기계학습 모델의 결과를 비교하여 최적의 성능을 발휘하는 모델을 선정하였다. 또한, 선정된 모델을 통해 기계학습 모델이 지반함몰 위험도 분류 시 활용한 영향인자의 중요도를 제시하고자 하였다.

## 2. 기계학습 알고리즘

### 2.1 Random Forest

Random Forest(RF) 알고리즘은 Breiman et al.(1984)이 제안한 회귀 및 분류 트리 기반의 앙상블(Ensemble) 모델이다(Lee et al., 2020). 앙상블 모델은 단일 알고리즘의 반복 또는 다수의 알고리즘을 학습하여 최적의 결과를 도출하는 방식으로, 단일 모델을 1회 학습한 결과보다 성능이 우수하다. RF는 다수의 Tree 알고리즘을 생성하여 각각의 Tree에서 도출된 결과를 토대로 가장 우수한 결과를 선정한다. 이와 같이 CART 기반으로 구성되어 있는 RF 알고리즘은 변수 선정에 자유롭고, Over Fitting의 위험도가 적은 특징을 갖고 있으며, 데이터 사이의 상관성이 높지 않아도 결과 도출에 있어 모델의 성능이 탁월한 장점이 있다(Park et al., 2019). 따라서 RF는 다양한 분야에서 기계학습 기법의 적용 시 회귀 및 분류문제 해결에 적극적으로 활용되고 있다.

### 2.2 XGBoost(eXtreme Gradient Boosting)

XGBoost 알고리즘은 선형 또는 Tree 기반 모델의 모델에서 나타나는 Over Fitting의 문제를 해결하기 위해 제시된 모델로, 대용량 데이터의 처리 및 학습 속도의 향상을 목적으로 개발되었다(Chen et al., 2016). XGBoost는 다수의 분류기를 생성하여 순차적으로 학습하고 각각의 모델에서 도출된 결과를 다음 모델에 반영하는 방법으로 문제를 해결하는 부스팅(Boosting) 기법이며, 주요 Hyper parameter는 Tree의 개수와 깊이 등이 있다.

### 2.3 LightGBM(Light Gradient Boosting Machine)

LightGBM은 Tree 알고리즘을 기반으로 한 고성능의 알고리즘이며, 회귀 및 분류 문제와 영향인자의 중요도 순위를 선정하는데 활용되고 있다. XGBoost와 유사한 Boosting 기법이 적용되고 있으며, 데이터의 일부를 활용하여 빠른 연산을 실시하고 특성을 감소시키는 방식으로 시간을 단축하는 특징을 갖고 있다(Lee et al., 2020). 따라서 LightGBM은 대용량 데이터를 빠른 속도로 처리하고, 높은 정확도를 나타내며, 활용된 영향인자 사이의 중요도를 도출할 수 있어 적극적으로 활용되고 있다.

### 2.4 AdaBoost(Adaptive Boosting)

AdaBoost 알고리즘은 전술한 XGBoost, LightGBM과 동일한 Boosting 기법이 적용된 알고리즘으로 약한 성능의 분류기를 선형으로 결합하여 하나의 강력한 분류기를 도출하는 알고리즘이다. AdaBoost는 분류기가 간단한 구조의 계층적인 체계를 이루고 있으며, 초기에는 모든 샘플에 대하여 동일한 가중치에 대한 학습을 실시한 후, 학습을 반복할수록 오류에 대한 가중치를 증가하는 방법으로 알고리즘의 성능을 향상시킨다(Lee, 2009).

### 2.5 모델의 평가지표

본 연구에서는 조건에 따라 구성된 데이터셋을 기계학습 모델에 적용하고 그 결과를 비교하고자 하였다. 모델의 성능 비교를 위한 평가 지표는 일반적으로 분류 모델의 평가지표로 활용되고 있는 정확도(Accuracy)와 F1-Score, AUC(Area Under the Curve)를 선정하였다.

정확도는 직관적으로 모델의 신뢰도를 평가할 수 있는 지표이나, 불균형한 데이터의 특징을 나타내는 데이터셋에 적용할 경우 명확한 모델의 평가가 어렵다. 따라서 본 연구에서의 정확도는 Train Data와 Test Data의 Score를 도출하여 모델의 과적합 여부를 확인하는데 사용되었으며, Score의 차이

가 적을수록 과적합의 위험이 적은 것으로 판단하는데 활용하였다.

F1-Score는 불균형한 데이터가 적용된 분류 모델의 객관적인 평가지표로 주로 활용되고 있으며, 모델이 True라고 예측한 데이터 중 실제 True인 데이터의 수(Precision)와 실제 True인 데이터에서 True라고 예측한 데이터의 수(Recall)의 조화 평균을 나타낸 지표이다(Lee et al., 2022). 이를 활용하면 예측 모델이 각각의 Class에 대해 적절하게 분류했는지에 대한 평가가 가능하다.

AUC는 Fig. 1에 나타난 ROC곡선의 면적을 통해 모델의 성능을 평가할 수 있는 지표이며, ROC곡선은 Recall과 Specificity를 이용하여 나타낸다. 모델의 세부 결과를 통해 도출된 Sensitivity와 Specificity 값을 활용하여 ROC 곡선을 그릴 수 있으며, 그려진 곡선의 면적 값을 AUC 값으로 활용한다. 일반적으로 AUC가 높을수록 모델의 신뢰도 및 성능이 우수하다고 평가된다. Table 1은 Fawcett(2005)가 제안한 AUC 값에 따른 모델의 성능을 나타내는 기준이며, 0.8 이상일

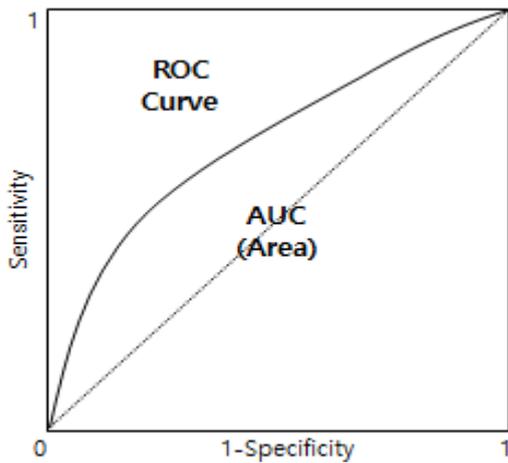


Fig. 1. Conceptual diagram of ROC curve

Table 1. Model evaluation according to AUC (Fawcett, 2005)

AUC	Evaluation
$AUC \geq 0.9$	Excellent
$0.8 \leq AUC < 0.9$	Good
$0.7 \leq AUC < 0.8$	Fair
$AUC < 0.7$	Poor

Table 2. Confusion Matrix

Confusion matrix		Prediction	
		Negative	Positive
Reference	Negative	True Negative (TN)	False Positive (FP)
	Positive	False Negative (FN)	True Positive (TP)

경우 모델의 성능이 우수하다고 평가할 수 있다.

Eq. (1)~(5)는 모델의 평가지표의 산출 방법을 나타낸 식이며, Table 2는 평가지표 산출을 위한 모델의 분류 결과를 나타낸 Confusion Matrix이다.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Recall(Sensitivity) = \frac{TP}{TP + FN} \quad (2)$$

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

$$Specificity = \frac{TN}{TN + FP} \quad (5)$$

### 3. 연구 방법 및 데이터셋 구축

#### 3.1 연구 방법 및 흐름

본 연구는 ○○시의 ○○구와 △△구를 대상지역으로 ArcGIS 프로그램을 활용하여 500m×500m 크기의 Grid로 분할하였다. Grid는 총 325개로 구분되었으며, 각 Grid 내에 포함되어 있는 6종 지하매설물을 1종의 지하매설물로 병합하여 속성정보와 밀집도를 추출하였고, 지반함몰 이력정보를 통해 Grid 내에 발생한 지반함몰 개수를 계산하여, 지반함몰 위험도를 산정했다.

위와 같은 방법으로 데이터셋을 구축한 뒤, 단위를 일정한 범위로 조정하는 표준화(Standardization)를 실시하고, 모델의 과적합 방지 및 평가를 위해 80:20의 비율로 훈련 데이터셋과 평가 데이터셋을 분할하였다. 본 연구에서 구축된 데이터셋은 지반함몰이 '0'개 발생한 수가 매우 많은 불균형한 데이터 특성을 나타내고 있다. 따라서 훈련 데이터셋에 소수의 데이터 특성을 부각하기 위한 Over Sampling 기법인 SMOTE를 적용하였다.

Sampling 기법이 적용된 훈련 데이터셋을 기계학습 알고리즘인 RF, XGBoost, LightGBM, AdaBoost에 적용하여 최적의 성능을 발휘하는 Hyperparameter를 조정하여 모델의 결과를 확인하였으며, 최적의 성능을 발휘하는 데이터셋의 형태와 기계학습 모델을 도출한 뒤, 영향인자의 중요도를 도출하였다. Fig. 2는 본 연구의 흐름도이다.

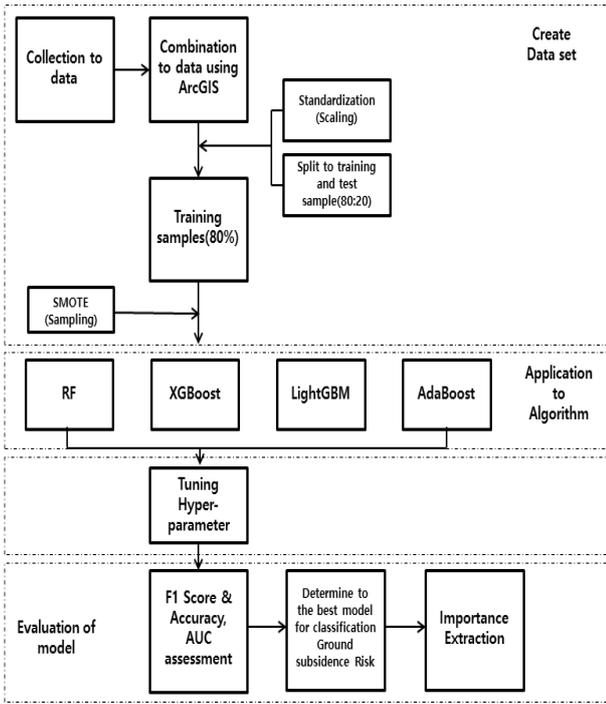


Fig. 2. Flow chart

### 3.2 대상지역의 구분 및 지하매설물 속성 데이터

지반함몰 위험도 예측을 위한 대상지역으로 ○○시의 ○○구와 △△구로 선정하였으며, 전술한 바와 같이 ArcGIS를 통해 대상지역을 500m×500m 크기의 325개 Grid로 구분하였다. 각 Grid 내에는 6종 지하매설물(상수, 하수, 전기, 가스, 난방, 통신관)의 속성 데이터와 지반함몰 이력 데이터가 포함되어 있으며, 6종 지하매설물을 1종의 지하매설물로 병합하여 속성 데이터를 추출하였다. 지하매설물 속성데이터에는 활용년수, 관중, 관경, 관길이, 매설심도, 경사도 등의 데이터가 포함되어 있으나, 결측값과 오류값이 다량 존재하여 실제 활용 가능데이터로는 활용년수, 관경, 관길이 있으며, 관로 전체의 밀집도를 계산하여 지반함몰 발생 영향인자로 활용하였다.

지반함몰 위험도 예측 모델의 성능 향상을 위해 영향인자의 범위를 구분하였다. 활용년수는 5년과 10년으로 구분하였으며, 관경은 50mm와 100mm로 구분하였고 사용된 단위로는 해당 범위에 속하는 관의 길이로 설정하였다. 예를 들어, Grid 내에 활용년수가 2년차인 지하매설물의 경우 활용년수 1~4년에 해당하는 Class에 관로의 길이를 반영하는 방법으로 설정하였으며, 데이터셋의 범주는 Table 3과 같다. Table 3에서 Year은 활용년수(5:5년 단위, 10:10년 단위), Diameter(50:50mm 단위, 100: 100mm 단위)는 관직경을 의미한다.

Table 3. Category of factors

Factors		Category
Year (year)	5	1~5, 6~10, 11~15, 16~20, 21~25, 26~30, 31~35, 36~40, 41~45, 46~50
	10	1~10, 11~20, 21~30, 31~40, 41~50
Diameter (mm)	50	1~50, 51~100, 101~150, 151~200, 201~250, 251~300, 301~350, 351~400, 401~450, 451~500, 501~550, 551~560
	100	1~100, 101~200, 201~300, 301~400, 401~500, 501~600

### 3.3 밀집도

선행 연구에 따르면 관로의 밀집도는 지반함몰과 유의미한 상관도를 보이는 것으로 나타났으며(Kim et al., 2021), 이에 따라 관로의 밀집도를 지반함몰 위험도 예측 모델의 영향인자로 활용하였다. 밀집도는 Grid 내에 관로를 대상으로 선형 밀도 분석을 실시하였으며, 단위 면적에 해당하는 관로의 길이를 계산하는 방법으로 밀집도를 산출하였다.

### 3.4 지반함몰 위험등급

데이터셋은 지하매설물의 활용년수와 관경, 밀집도, 지반함몰 위험등급으로 구성되어 있으며, 지반함몰 위험등급은 Grid 내에 발생한 지반함몰 개수를 파악하여 모델이 최적의 성능을 발휘할 수 있는 개수에 따른 등급을 임의로 3등급으로 선정하였다.

Grid 내에 발생한 지반함몰은 0~43개의 범위로 발생하였으며, 전체 중 약 49%가 지반함몰 발생 횟수 '0'회로 나타났다. 지반함몰 위험등급은 3등급(0~2등급)으로 결정하였으며, 0등급(약 49%)은 지반함몰 발생횟수가 0회, 1등급(약 24%)은 1~2회, 2등급(약 27%)은 3회 이상으로 구분하였다.

기계학습에 적용하기 위한 데이터셋은 지반함몰 위험도 '0'등급의 비율이 매우 높은 불균형한 특징을 갖고 있다. 따라서 이를 해소하기 위해 SMOTE Sampling 기법을 Train set에 적용하였으며, 위험도 '1'과 '2'의 데이터수가 '0'과 동일한 127개로 맞추어졌다.

### 3.5 데이터셋 조건

최적의 성능을 발휘하는 지반함몰 위험도 예측 모델을 선정하기 위해 지하매설물 속성의 범주에 따른 4개 조건의 데이터셋을 기계학습 모델에 적용하였다. 데이터셋의 조건은 Table 4와 같으며, 지반함몰 위험등급과 밀집도는 고정하였다.

Table 4. Configuration of datasets

	Year (year)	Diameter (mm)	Density	Risk level
MD-1	10	100	Density	3
MD-2	10	50		
MD-3	5	100		
MD-4	5	50		

## 4. 기계학습 모델의 적용 방법 및 결과

### 4.1 기계학습 모델의 적용 방법

지반함몰 위험도 예측 모델을 구현하기 위해 Python 3.8과 전처리 및 기계학습 알고리즘 패키지가 포함되어 있는 Scikit-learn 라이브러리를 활용하였다. 해당 패키지를 활용해 4가지 조건의 데이터셋을 RF, XGBoost, LightGBM, AdaBoost 알고리즘에 적용하여 지반함몰 위험도 예측 모델을 제시하고자 하였다. 각각의 데이터셋 적용 시 최적의 결과를 도출할 수 있도록 Hyperparameter를 튜닝하였으며, Table 5는 최적의 결과를 도출한 Hyperparameter를 나타낸 표이다.

Hyperparameter의 튜닝은 시행착오 방법으로 최적의 결과가 도출되고 과적합을 회피하는 매개변수의 값으로 선정하였으며, Hyperparameter에서 Estimators는 단일 모델(Tree)의 개수, Max Depth는 Tree의 깊이, Learning rate는 학습률을 의미한다. 위의 Hyperparameter의 값이 커질수록 모델의 성능은 향상되지만, 과적합의 위험이 급격히 상승하므로 적절한 수준을 유지하는 것이 중요하다.

Table 5. Summary of hyper parameters in the model

Model	Hyper parameter
RF	Estimators(500) Max depth(4)
XGB	Estimators(500) Max depth(3) Learning rate(0.006)
LightGBM	Estimators(200) Max depth(3) Learning rate(0.01)
AdaBoost	Estimators(100) Learning rate(0.05)

### 4.2 기계학습 모델 비교 및 선정

지반함몰 위험도 예측 모델의 결과를 Table 6에 나타냈다. 평가지표로는 과적합 여부를 확인할 수 있는 Test, Train Score와 모델의 성능을 판단하는 F-1 Score이며, micro는 전체 클래스의 F-1 Score 평균 값을 의미한다.

모델의 평가지표를 비교한 결과, 활용년수 10년 단위, 관

Table 6. Result comparison for each models

	Model	Test Score	Train Score	F-1 Score (micro)
MD-1	RF	0.600	0.691	0.580
	XGB	0.615	0.675	0.610
	LightGBM	0.600	0.669	0.600
MD-2	Ada	0.600	0.715	0.620
	RF	0.708	0.808	0.730
	XGB	0.800	0.871	0.800
MD-3	LightGBM	0.769	0.850	0.770
	Ada	0.523	0.675	0.500
	RF	0.554	0.669	0.540
MD-4	XGB	0.554	0.653	0.570
	LightGBM	0.600	0.713	0.620
	Ada	0.554	0.607	0.570
	RF	0.692	0.764	0.700
MD-4	XGB	0.692	0.798	0.690
	LightGBM	0.646	0.701	0.660
	Ada	0.723	0.772	0.730

경 50mm로 구분한 데이터셋에서 상대적으로 우수한 F-1 Score가 도출된 것으로 나타났다. 특히, MD-2의 XGB 모델의 F-1-Score가 가장 우수했으며, Test와 Train Score의 차이가 0.071로 나타났다. Test와 Train Score를 활용한 과적합 여부 판단 시 명확한 기준이 없으므로 본 연구에서는 0.1 이하로 차이가 나면 과적합을 회피한 것으로 판단하였다. 따라서 해당 모델은 과적합을 회피한 것으로 판단된다. 이는, 밀집도가 제외된 지하매설물 속성을 활용한 지반함몰 위험도 예측 모델(Lee, 2022)의 결과(F1-Score:0.770) 보다 우수한 분류 성능을 발휘하는 것으로 나타났다. 또한, SMOTE Sampling 기법을 적용하지 않은 MD-2 XGB모델의 F-1 Score는 0.750으로 나타나, Sampling 기법의 적용이 적절한 것으로 판단된다.

MD-1과 MD-3, MD-2와 MD-4의 비교를 통해 활용년수 구분에 따른 모델의 성능을 비교한 결과, 활용년수 구분에 따라 모델의 성능 차이는 경향을 찾기 어려웠고, MD-1과 MD-2, MD-3와 MD-4의 비교를 통해 관경 구분에 따른 모델의 성능을 비교한 결과, 관경 50mm로 구분한 데이터셋을 적용한 모델의 성능이 상대적으로 우수한 평가지표를 도출하였다.

따라서 ○○시의 ○○구와 △△구의 가장 효율적인 지반함몰 위험도 예측 모델은 MD-2 XGB로 선정되었으며, 해당 모델의 신뢰도를 확인하기 위해 Fig. 3에 ROC곡선을 나타냈다. 위험도 0과 2의 AUC는 0.9 이상으로 매우 좋은 성능을 발휘하고, 위험도 1은 0.85로 좋은 분류 성능을 발휘하는 것으로 나타났다.

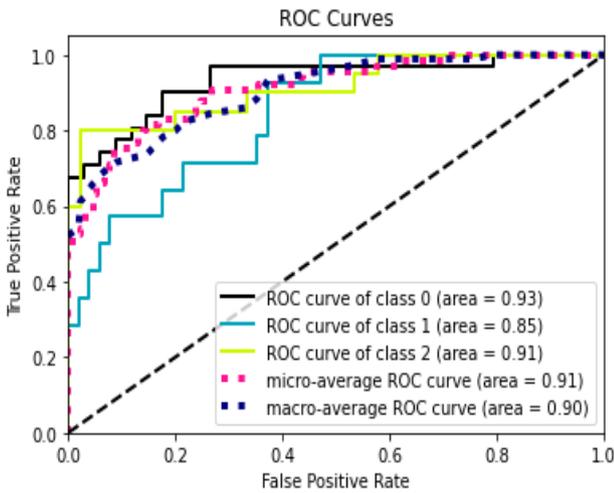


Fig. 3. ROC curve of MD-2 XGB model

### 4.3 선정 모델을 활용한 영향인자 중요도

본 연구 대상지역에서의 최적 지반함몰 위험도 예측 모델은 MD-2 XGB로 선정되었으며, XGB 알고리즘은 해당 모델이 문제를 해결할 때 활용한 영향인자의 중요도를 도출하는 기능이 포함되어 있다. 이러한 기능을 활용하여 지반함몰 위험도 예측 시 활용한 영향인자(활용년수, 관경, 밀집도)의 중요도를 Fig. 4에 도출하였다.

중요도 도출 결과, 활용년수 40~50년(Y50) 사이의 관이 지반함몰에 미치는 영향이 큰 것으로 나타났으며, 관직경 50mm 이하(DTR\_50)의 관, 밀집도(Density) 순으로 나타났다. 또한, 관직경 200mm(DTR\_200) 이상의 관과 활용년수 10년 이하(Y10)의 관의 지반함몰 발생 영향도는 낮은 것으로 나타났다.

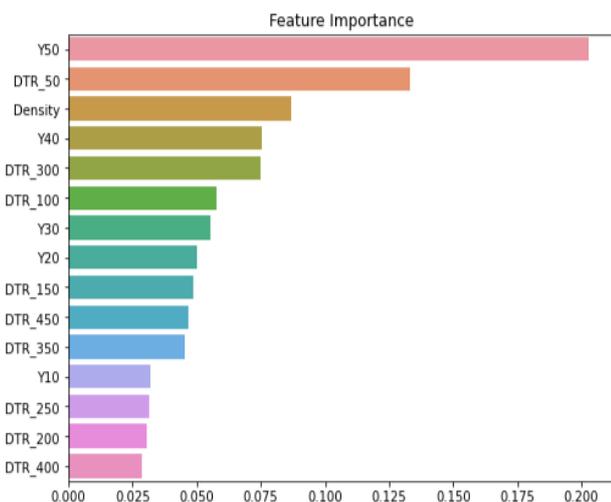


Fig. 4. Feature importance of MD-2 XGB model

## 5. 결 론

본 연구는 ○○시 2개구의 지하매설물 속성인 활용년수, 관직경과 관로 밀집도, 지반함몰 위험도를 활용하여 대상 지역의 최적 지반함몰 위험도 예측 모델을 제시하고자 하였다. 이를 위해 ArcGIS 프로그램을 활용하여 500m×500m 크기로 Grid를 분할하여, Grid 내 지하매설물 속성 데이터와 지반함몰 이력 데이터를 통해 데이터셋을 구축하였고, 전처리 과정을 거쳐 기계학습 모델에 적용하여 최적의 성능을 발휘하는 모델을 선정하였다. 또한 선정된 모델의 ROC 곡선을 통해 AUC 값을 확인하였으며, 활용인자의 중요도를 도출하여 확인하였다. 본 연구의 요약과 결론은 다음과 같다.

- 시 2개구의 최적 지반함몰 위험도 예측 모델을 위한 데이터셋은 활용년수는 10년 단위, 관경은 5년 단위로 구성하는 것이 모델의 성능이 가장 우수하게 발휘되는 것으로 나타났다.
- 기계학습 알고리즘(RF, XGB, LightGBM, Ada)에 적용한 결과, MD-2 XGB 모델에서 상대적으로 높은 F-1 Score가 도출되었고, Test와 Train의 Score 차가 크지 않아 과적합도 회피한 것으로 나타났다. 또한, AUC도 0.8을 상회하여 해당 모델을 가장 적절한 모델로 선정하였다.
- 지반함몰 영향인자로 밀집도가 추가된 MD-2 XGB와 밀집도가 제외된 모델과의 평가지표를 비교하였을 때, MD-2 XGB 모델의 F-1 Score가 높게 나타났으므로, 지반함몰 위험도 예측 분석 시 밀집도를 추가해 분석하는 것이 바람직할 것으로 판단된다.
- MD-2 XGB 모델을 통해 지반함몰 위험도 예측 분석 시 활용된 영향인자의 중요도를 도출한 결과, 지하매설물의 활용년수가 40년 이상 된 관이 지반함몰 위험도 예측에 가장 크게 활용되었고, 관직경 50mm 이하 관, 밀집도의 순으로 나타났다. 또한, 관직경 200mm 이상 관 및 활용년수 10년 이하의 관은 지반함몰 위험도 예측 시 중요도가 떨어지는 것으로 나타났다.

본 연구에서 제안된 지반함몰 위험도 예측 모델 구축 방안 및 모델을 통해 추후 광범위한 지역의 지반함몰 위험도 예측 모델 제안 시 방법 및 방향을 제시할 수 있을 것으로 기대된다. 또한, 모델을 통해 나타난 위험지역을 GPR 탐사 시 우선순위 대상으로 선정하여 효율적인 탐사가 이루어질 수 있을 것으로 기대된다.

## 감사의 글

본 연구는 (23주요-대1-임무)지하 공간 정보 정확도 개선 및 매설관 안전관리 기술개발 (4/4) 지원으로 수행되었으며, 이에 깊은 감사를 드립니다.

## References

1. Breiman, L., Friedman, J., Stone, C. and Olshen, R. (1984), Classification and regression trees, Taylor & Francis.
2. Chen, T. and Guestrin, C. (2016), XGBoost: A Scalable Tree Boosting System, KDD'16 Proceedings of the 22<sup>nd</sup> ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785~794.
3. Han, M. S. (2017), A risk assessment of ground subsidence by GPR and CCTV investigation, Master's thesis, Seoul National University of Science and Technology (In Korean).
4. Jin, Y. S. (2018), The Analysis on Correlation of Precipitation and Risk Factors to the Soil Subsidence, Ph D. dissertation, Chonnam National University, pp. 104~105 (In Korean).
5. Kim, J. Y., Kang, J. M., Choi, C. H. and Park, D. H. (2017), Correlation analysis of sewer integrity and ground subsidence, Journal of the Korean Geo-Environmental Society, Vol. 18, No.6, pp. 31~37 (In Korean).
6. Kim, J. Y., Kang, J. M. and Choi, C. H. (2021), Correlation analysis of the occurrence of ground subsidence according to the density of underground pipelines, Journal of the Korean Geo-Environmental Society, Vol. 22, No. 11, pp. 23~29 (In Korean).
7. Kim, K. Y. (2018), Susceptibility Model for Sinkholes Caused by Damaged Sewer Pipes Based on Logistic Regression, Master's thesis, Seoul National University (In Korean).
8. Kuwano, R., Horii, T., Kohashi, H. and Yamauchi, K. (2006), Defects of sewer pipes causing cave-in's in the road, Proc. 5th International Symposium on New Technologies for Urban Safety of Mega Cities in Asia, Phuket, Thailand, pp. 347~353.
9. Lee, J. T. (2009), A Study on Multi-Class Classification Method Using Adaboost, Master's thesis, Hanyang University (In Korean).
10. Lee, S. H., Yoon, Y. A., Jung, J. H., Sim, H. S., Chang, T. W. and Kim, Y. S. (2020), A machine learning model for predicting silica concentrations through time series analysis of mining data, Journal of Korean Society for Quality Management, Vol. 48, No. 3, pp. 511~520 (In Korean).
11. Lee, S. Y., Kang, J. M. and Kim, J. Y. (2022), Development of machine learning model to predict the ground subsidence risk grade according to the Characteristics of underground facility, Journal of the Korean Geo-Environmental Society, Vol. 23, No. 4, pp. 5~10 (In Korean).
12. Lee, S. Y., Kim, J. Y., K., J. M. and Baek, W. J. (2022), Comparison of machine learning models to predict the occurrence of ground subsidence according to the characteristics of sewer, Journal of Korean Geo-Environmental Society, Vol. 23, Issue. 4, pp. 5~10 (In Korean).
13. Mukunoki, T., Kuwano, N., Otani, J. and Kuwano, R. (2009), Visualization of three dimensional failure in sand due to water inflow and soil drainage from defected underground pipe using X-ray CT, Soils and Foundations, Vol. 49, No. 6.
14. Park, E. J., Park, J. H. and Kim, H. H. (2019), Mapping species-specific optimal plantation sites using random forest in Gyeongsangnam-do province, South Korea, Journal of Agriculture & Life Science, Vol. 53, No. 6, pp. 65~74 (In Korean).
15. Seoul Seokchon-dong Cavity Cause Investigation Committee (2014), Cause Analysis of Cavity at Seokchon Underground Roadway and Road Cavity Special Management Measures.
16. Takeuchi, D., Fukatani, W., Miyamoto, T. and Yokota, T. (2017), Using decision tree analysis to extract factors affecting road subsidence, Journal of the Japan sewage works association, Vol. 54, No. 657, pp. 124~133.
17. Tom Fawcett (2005), An introduction to ROC analysis, Patter Recognition Letters, Edited by Francesco Tortorella, Vol. 27 Issue 8, pp. 861~874.