The Journal of Engineering Geology
pISSN : 1226-5268 eISSN : 2287-7169

Vol. 30, No. 4, December, 2020, pp. 649-661
https://doi.org/10.9720/kseg.2020.4.649

RESEARCH ARTICLE

로지스틱 회귀분석을 이용한 도로비탈면관리시스템 데이터 활용 검토 연구

우용훈[1] · 김승현[2] · 양인철[2] · 이세혁[1*]

[1]한국건설기술연구원 전임연구원, [2]한국건설기술연구원 연구위원

The Study for Utilizing Data of Cut-Slope Management System by Using Logistic Regression

Yonghoon Woo[1] · Seung-Hyun Kim[2] · Inchul Yang[2] · Se-Hyeok Lee[1*]

[1]Research Specialist, Korea Institute of Civil Engineering and Building Technology
[2]Research Fellow, Korea Institute of Civil Engineering and Building Technology

Abstract

Cut-slope management system (CSMS) has been investigated all slopes on the road of the whole country to evaluate risk rating of each slope. Based on this evaluation, the decision-making for maintenance can be conducted, and this procedure will be helpful to establish a consistent and efficient policy of safe road. CSMS has updated the database of all slopes annually, and this database is constructed based on a basic and detailed investigation. In the database, there are two type of data: first one is an objective data such as slopes' location, height, width, length, and information about underground and bedrock, etc; second one is subjective data, which is decided by experts based on those objective data, e.g., degree of emergency and risk, maintenance solution, etc. The purpose of this study is identifying an data application plan to utilize those CSMS data. For this purpose, logistic regression, which is a basic machine-learning method to construct a prediction model, is performed to predict a judging-type variable (i.e., subjective data) based on objective data. The constructed logistic model shows the accurate prediction, and this model can be used to judge a priority of slopes for detailed investigation. Also, it is anticipated that the prediction model can filter unusual data by comparing with a prediction value.

Keywords: cut-slope management system, machine-learning, logistic regression, prediction model

***Corresponding author:** Se-Hyeok Lee
E-mail: sehyeoklee@kict.re.kr

Received: 24 November, 2020
Revised: 9 December, 2020
Accepted: 9 December, 2020

초 록

도로비탈면관리시스템은 전국 도로 비탈면 현황을 파악하고 위험등급을 산정하여 유지대책 선정 및 사전에 비탈면 붕괴를 차단하여 국민의 안전을 도모하기 위해 만들어졌다. 이를 위해 전국 국도에 위치한 깎기비탈면에 대해 기초 · 정밀조사를 수행하여 데이터베이스를 구축하고 매년 갱신되고 있다. 수집된 데이터는 수치형과 문자형으로 구성되어 있으며, 사면에 대한 객관적인 정보와 전문가의 판단에 의해 결정된 주관적인 정보로 구성되어 있다. 본 연구에서는 도로비탈면관리시스템에서 관리하는 데이터 활용 가능성을 검토하기 위해, 기계학습인 로지스틱 회귀분석을 이용하여 독립적인 정보를 이용한 주관적 정보 예측 모델을 구축하고 검증하였다. 수행결과, 구축된 확률모델을 이용하여 높은 정확도로 주관적 판단이 필요한 정보들을 예측할 수 있음을 확인하였다. 또한, 구축된 모델을 활용하여 새로 수집된 정보와 모델로부터의 예측값을 비교 · 검토를 통해 고품질의 데이터를 구축할 수 있을 것으로 기대된다.

주요어: 도로비탈면관리시스템, 기계학습, 로지스틱 회귀분석, 예측모델

서론

도로비탈면관리시스템(Cut Slope Management System, CSMS)은 국토교통부 위탁고시에 따라 한국건설기술연구원과 한국시설안전공단이 운영하고 있으며, 이들 기관은 지역 별로 관할 도로를 나누어 비탈면 관리 위탁업무를 수행하고 있다(KICT, 2019). CSMS 운영은 전국 도로 비탈면 현황을 파악하고 위험등급을 산정한 후 연차별로 효율적인 비탈면 유지대책을 도출하여, 사전에 비탈면 붕괴를 차단함으로써 도로를 이용하는 국민의 안전도모를 목표로 한다(KICT, 2019). CSMS는 국도 주변의 모든 깎기비탈면에 대한 데이터베이스를 구축하고 있으며, 매년 이를 갱신하고 있다. 이렇게 구축된 데이터를 바탕으로 위험성이 높은 비탈면에 대해 전문가에 의한 정밀조사가 수행되며 적정대책 공법 안이 마련 · 제시된다. 또한, 이러한 의사결정 과정을 기반으로 투자 우선순위가 결정되어 국가 예산의 투명하고 효율적인 집행이 이루어지게 된다. 최근 CSMS 업무 중 하나로, 스마트폰 · 웹기반을 이용하여 현장에서 바로 사용가능한 이동식 데이터베이스 체계 구축 연구가 진행 중이며, 이를 통해 더욱 빠른 데이터 수집이 가능할 것이다. 따라서 수집되는 데이터의 품질 관리와 구축된 데이터 활용 방안 고안은 향후 필수적인 연구 분야이다(KICT, 2019).

CSMS의 주 데이터는 크게 두 가지로 분류될 수 있다. 첫 번째는 기초조사 자료이며, 두 번째는 이를 바탕으로 수행되는 정밀조사에서 수집되는 자료이다. 기초조사는 국도의 확장, 신설노선 · 우회도로 개통, 위험도로 선형개량, 승격국도 발생에 의해 새로운 비탈면이 준공되었을 때, 국토관리사무소(또는 지자체)로부터 행정구역, 구간, 연장 등 기초사항 정보를 받고, 비탈면 기초조사 매뉴얼에 따라 비탈면 기본정보를 직접 현장에서 취득하는 것을 말한다. 이러한 기초조사 자료는 위험한 깍기 비탈면을 파악할 수 있는 기본 정보를 내포하고 있으며, 정밀조사 우선순위를 결정하는 주요한 자료이다. 또한, 도로 2종시설물을 파악하거나 상시계측 시스템 설치 결정 등 CSMS의 연구 업무를 위한 자료로 활용되며, 특히 시특법(시설물의 안전 및 유지관리에 관한 특별법) 변경에 따른 신규 2종 시설물 파악에 매우 유용한 자료이다.

반면, 기초조사 이후에 수행되는 정밀 조사는 대상 위험 비탈면의 불연속면 특성, 풍화도, 암반강도, 누수 현황, 배수시설 상태 등을 정밀하고 상세하게 조사하는 것으로, 수집된 정밀조사 자료는 안정성 해석을 위한 기본 자료 및 대책공법 수립 등에 사용된다. 특히, 2016년에 시특법이 개정됨에 따라 2종 시설물인 비탈면 수가 증가하였으며, 신규로 편입된 비탈면은 민간 기업에 의해 정밀조사가 새로이 수행되었다. 그런데 기초조사와 달리 정밀조사는 전문가에 의해 수행되기 때문에 전문가의 판단이 매우 중요하다. 이 과정에서 조사 수행자의 전문성 부족 및 자의적 해석 등으로 인해 잘못된 대책 공법 설계 사례가 발생할 수 있다. 이러한 인간 오차(Human Error)로 인해, 실제 대책공법이 필요한 위험 사면이 방치되어 인명 및 경제적 피해가 초래될 수 있으며, 또는 불필요한 대책공법 수행으로 비경제적 문제가 발생할 수 있다.

본 연구는 전국 비탈면 기초조사 및 정밀조사를 통해 수집된 객관적(정보형) 및 주관적(판단형) 데이터들 중 일부를 선정하고 로지스틱 회귀분석을 이용하여 예측 모델을 구축하였다. 이때, 구축된 모델의 정확도를 검토하고 이를 활용하여 누락 데이터 예측 혹은 고품질 데이터 구축 등 기계학습을 이용한 데이터 활용 가능성 파악을 목표로 한다. 회귀분석에 사용된 CSMS 데이터는 2006년부터 2019년에 수집된 비탈면 기초 · 정밀조사 자료이며, 이 자료는 수치형 데이터와 더불어 문자형 데이터로 구성되어 있다. 기초조사 자료는 대게 객관적인 정보인 반면에, 정밀조사 자료는 주관적인 판단형 데이터를 포함한다. 로지스틱 회귀분석을 수행하여 구축된 확률 모델은 기초조사 자료의 객관적인 정보형 항목들을 바탕으로 정밀조사의 판단형 항목을 예측할 수 있으며, 이를 통해 전문가의 의사결정 수행에 있어 인간 오차를 줄이는데 활용될 수 있다. 또한, 사전정보를 통해 대상 사건의 확률을 예측할 수 있다는 '베이지안(Bayesian)' 관점에서, 구축된 로지스틱 모델은 사전 정보인 기존 데이터를 통해 신규 비탈면 혹은 대책공법이 필요한 사면의 판단형 항목에 대해 예측 및 확률 추론이 가능하다. 이와 같이 정밀조사 이전에 각 사면에 대해 판단형 항목들을 사전에 예측함으로써 과학적인 예산 투자 우선순위 결정이 가능하며 기존 업무의 효율성 증대를 도모할 수 있을 것으로 기대된다.

CSMS 기초조사 및 정밀조사 데이터

CSMS 기초조사

기초조사는 전국 국도변에 있는 모든 깎기비탈면의 기초적인 속성 항목(관리 및 위치정보, 일반현황, 비탈면 특성, 조사자 소견, 사진 등)을 파악하고 관련 자료를 취득하는 전수조사를 의미한다(KICT, 2019). 취득한 자료는 Fig. 1과 같이 체계적·과학적인 비탈면 관리를 위해 각 비탈면에 고유코드를 부여하여 데이터베이스화되며, 조사우선순위를 결정하는 주요 자료가 된다. 2001년에 도로관리기관이 기초조사를 수행하였지만, 그 후 국도신설·위험도로개량으로 인해 비탈면 개소수가 급격히 증가함에 따라 2006년부터 연구기관들에 위임되어 수행되고 있다. 현재 국도 내 전 노선에 대한 기초조사는 완료된 상태로, 매년 국도 확·포장 및 개량 공사 등으로 새로이 조성된 비탈면에 대해 수행되고 있다.

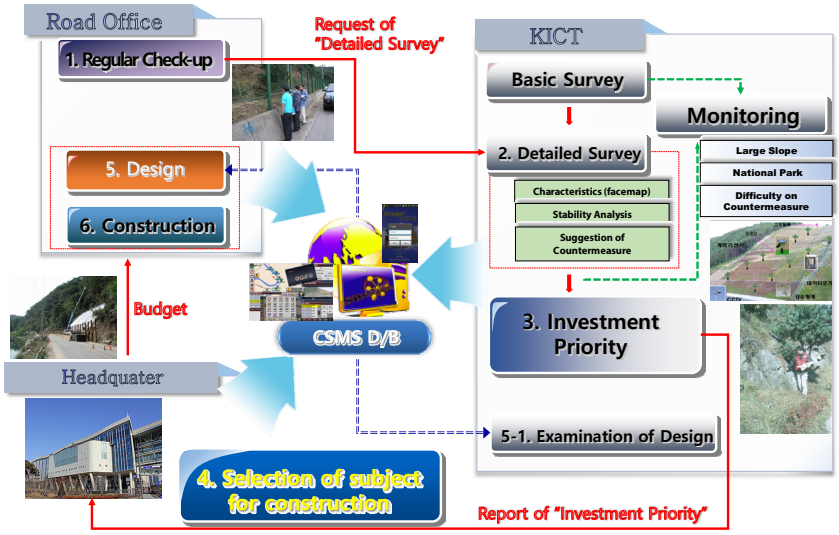


Fig. 1. Flowchart of cut-slope management system.

CSMS 정밀조사

정밀조사는 기초조사 자료를 바탕으로 선정된 위험등급 사면에 대해 비탈면 전문가가 수행하는 상세조사로, 1) 대상 비탈면의 위치, 연장, 높이, 경사, 등 규모와 관련된 정보, 2) 구성 재료의 종류, 위험구간 비율, 토층심도, 지하수 현황, 계곡부 유무, 토질 종류, 불연속면 특성·방향·종류, 풍화도, 암반강도, 붕괴발생이력 등 붕괴와 관련된 정보, 3) 적용 대책공법의 현황, 이격거리, 필요공법의 중요도 안정화와 관련된 정보로 구성되어 있다(KICT, 2019). 정밀조사는 크게 해빙기 정밀조사, 수시 정밀조사, 일반 정밀조사, 기타(위험그룹) 정밀조사로 구분되고, 이러한 정밀조사들을 근거로 구간별 특성이 고려된 현황도가 작성되어 위험구간에 대해 최적의 대책공법이 선정된다.

사용된 CSMS 데이터 개요

본 연구에서 사용된 CSMS 데이터는 2006~2019년에 수집된 30,751개 사면에 대한 조사 자료이며, Fig. 2와 같이 사면 기본 정보 및 특성과 같은 객관적인 정보와, 위험도 및 대책공법과 같은 전문가에 의한 판단형 정보로 이루어져 있다. 객관적인 정보형 데이터 예로는 '시/도/군/구', '왕복/편도', '차선수', '조사년도', '조사/미조사', '길이', '최대높이', '각도', '구배', '상부경사', '이격거리', '소단개소', '사면종류', '주변지형', '지하수', '누수위치 세로', '누수위치 가로', '풍화도',

'불연속면 방향성', '사면형상', '측면형상', '계곡부', '붕괴이력', '뜬돌', '낙석', '암종', '토층심도', '암반형태', '불연속면' 등이 있으며, 주관적인 판단형 데이터로는 '계측추천', '위험도', '피해도', '붕괴유형', '등급재조정', '필요 주공법 종류', '조치' 등이 있다. 다음에 설명될 로지스틱 회귀분석에서 앞서 언급된 29개의 정보형 데이터 항목들을 이용하여 판단형 데이터 중 해당 사면의 응급·미응급을 의미하는 조치 항목에 대한 모델을 구축하였다.

No.	절토사면 코드			조사기관	과대	최대높이	경사		상부공사	마격거리	소단계소	사면종류	주변지형	지하수	누수위치		풍화도	불연속면방향성	사면형상	측면형상	계곡부	붕괴이력	뜬돌	낙석	암종	토층심도(m)	암반형태	불연속면	계측추천	2중시설물	위험도	피해도	붕괴유형	위험등급	등급재조정
	년도 국도,호선	상/하	번호				각도	구배							세로	가로																			
1	07WK06	D	00100	KISTEC	150	16	45	1	-10	2	2	혼합	산악	damp	중	중	삼한풍화	확인불가	쌍형	직선형	0	무	무	무	화강암	6	irregular	갈리	비추천	부	하	하	무	C	II
2	07WK06	D	00200	KISTEC	50	13	50	0.8	5	2	0	토사	산악	wet	중	중	삼한풍화	확인불가	요형	직선형	0	유(미영향)	무	무	화강암	10	crushed	무	비추천	부	하	하	표층	C	III
3	07WK06	D	00300	KISTEC	73	13	52	0.8	10	1.5	0	혼합	산악	damp	중	중	삼한풍화	확인불가	쌍형	직선형	0	무	무	무	화강암	4	irregular	갈리	비추천	부	하	하	무	C	II
4	07WK06	D	00400	KISTEC	74	12	45	1	30	1.5	0	혼합	산악	wet	하	우	삼한풍화	확인불가	직선형	직선형	0	무	무	무	화강암	3	irregular	갈리	비추천	부	하	하	무	C	II
5	07WK06	D	00500	KISTEC	64	17	45	1	23	1	3	토사	산악	wet	중	중	완전풍화	확인불가	쌍형	직선형	0	무	무	무	화강암	15	crushed	무	비추천	부	하	하	무	C	II
6	07WK06	D	00600	KISTEC	50	17	55	0.7	30	1.5	0	혼합	산악	wet	상	좌	삼한풍화	확인불가	직선형	직선형	1	유(영향)	무	무	화강암	5	irregular	갈리	비추천	부	중	중	표층	C	III
7	07WK06	D	00700	KISTEC	70	13	55	0.7	26	1	0	혼합	산악	damp	상	우	삼한풍화	확인불가	쌍형	직선형	0	유(영향)	무	무	화강암	3	irregular	갈리	비추천	부	중	하	표층	C	III
8	07WK06	D	00800	KISTEC	35	11	55	0.7	25	4	0	토사	산악	damp	중	중	완전풍화	확인불가	요형	직선형	0	유(미영향)	무	무	화강암	10	crushed	무	비추천	부	하	하	표층	C	III
9	07WK06	D	00900	KISTEC	20	8	53	0.8	-15	1.5	0	혼합	산악	damp	중	중	삼한풍화	확인불가	직선형	직선형	0	무	무	무	화강암	4	irregular	갈리	비추천	부	하	하	무	B	II
10	07WK06	D	01000	KISTEC	48	13	45	1	-40	1	2	토사	산악	damp	중	중	완전풍화	확인불가	요형	직선형	0	무	무	무	화강암	10	crushed	무	비추천	부	하	하	무	C	II
11	07WK06	D	01100	KISTEC	40	11	50	0.8	18	1	0	혼합	산악	damp	중	중	삼한풍화	확인불가	직선형	직선형	0	무	무	무	화강암	5	irregular	갈리	비추천	부	하	하	무	C	II
12	07WK06	D	01200	KISTEC	90	20	40	1.2	5	1.5	3	토사	산악	damp	중	중	완전풍화	확인불가	쌍형	직선형	0	무	무	무	화강암	12	crushed	무	비추천	부	하	중	무	C	II
13	07WK06	D	01300	KISTEC	50	23	40	1.2	10	3	3	혼합	산악	wet	상	중	삼한풍화	확인불가	요형	직선형	0	무	무	무	화강암	10	irregular	갈리	비추천	부	중	중	무	D	II
14	07WK06	D	01400	KISTEC	40	8	45	1	-10	1	0	혼합	산악	damp	중	중	삼한풍화	확인불가	쌍형	직선형	0	유(영향)	무	무	화강암	4	irregular	갈리	비추천	부	하	하	무	C	III
15	07WK06	D	01500	KISTEC	50	16	52	0.8	15	1.5	0	혼합	산악	damp	상	중	삼한풍화	확인불가	쌍형	직선형	0	유(미영향)	소	소	화강암	7	irregular	갈리	비추천	부	중	중	표층	C	II
16	07WK06	D	01600	KISTEC	40	11	61	0.6	22	1.5	1	암반	산악	damp	중	좌	삼한풍화	일치	직선형	요형형	0	낙석	소	소	화강암	1	irregular	갈리	비추천	부	중	중	표층	C	I
17	07WK06	D	01700	KISTEC	45	10	60	0.6	30	1.5	0	암반	산악	damp	중	중	보통풍화	일치	직선형	요형형	0	유(영향)	중	중	화강암	3	irregular	갈리	비추천	부	중	중	표층	C	III
18	07WK06	D	01800	KISTEC	40	21	40	1.2	-15	1.5	3	혼합	산악	damp	상	중	삼한풍화	확인불가	쌍형	직선형	0	무	무	무	화강암	10	irregular	갈리	비추천	부	하	중	무	C	II
19	07WK06	D	01900	KISTEC	100	18	40	1.2	10	1	3	혼합	산악	damp	상	중	삼한풍화	확인불가	쌍형	직선형	0	유(미영향)	무	무	화강암	10	irregular	갈리	비추천	부	하	중	표층	C	III
20	07WK06	D	02000	KISTEC	90	17	40	1.2	10	1.5	3	토사	산악	damp	상	중	완전풍화	확인불가	쌍형	직선형	0	유(미영향)	무	무	화강암	0	crushed	무	비추천	부	하	중	표층	C	III
21	07WK06	U	02100	KISTEC	45	9	45	1	-60	1.5	1	혼합	산악	damp	중	중	삼한풍화	확인불가	직선형	직선형	0	무	무	무	화강암	4	irregular	갈리	비추천	부	하	하	무	B	II

Fig. 2. Some part of cut-slope management system data.

로지스틱 회귀분석 및 데이터 전처리

로지스틱 회귀분석(Logistic Regression)

로지스틱 회귀분석은 기계학습의 일종으로 0과 1처럼 두 클래스(Binary)로 이루어진 종속변수와 독립변수간의 관계를 구체적인 함수로 나타내는 확률 모델로서, 구축된 확률모델은 독립변수의 선형 결합을 이용하여 사건(종속변수)의 발생 가능성 예측이 가능하다(Lee and Kim, 2004; Quan et al., 2011; Yeon, 2011; Lee and Kim 2012; Chae et al., 2004; Baek et al., 2016).

$$odds = \frac{p(y=1|x)}{1-p(y=1|x)} \tag{1}$$

$$\text{logit}(p) = \log\frac{p}{1-p} \tag{2}$$

$$y = \frac{\exp(\sum b_0 + b_i x_i)}{1+\exp(\sum b_0 + b_i x_i)} \tag{3}$$

식 (1)에서 $p(y=1|x)$은 조건부 x 변수에 대해 1이라는 y사건이 발생할 확률을 의미한다. 로지스틱 회귀분석은 식 (1)과 같이 성공할 확률과 실패할 확률의 비를 타나내는 오즈(*Odds*) 개념을 이용한다. 그리고 식 (2)는 오즈의 로그값으로 로짓 변환($\text{logit}(\bullet)$) 을 의미한다. 로짓 변환 값은 0과 1사이의 종속변수의 확률값을 오즈 개념을 이용하여 음의 무한대에서 양의 무한대로 변환해주며, 이러한 정의들 덕분에 로지스틱 회귀분석은 두 클래스로 표현되는 종속변수에 대한 확률모

델을 구축할 수 있게 된다. 최종적으로, 종속변수(y)와 독립변수(x)간의 선형관계를 산정된 계수 b_i를 이용하여 식 (3)과 같이 확률모델로 나타낼 수 있다.

로지스틱 회귀분석은 응급, 미응급과 같이 0과 1로 표현될 수 있는 문자형 데이터에 대해 적용 가능하다는 장점이 있으며, 일반적인 회귀분석과는 달리 범주형 종속변수를 대상으로 하며 입력 데이터(독립변수의 선형결합)에 대해 결과가 분리되기 때문에 분류 기법으로도 알려져 있다. 본 논문에서는 로지스틱 회귀분석을 수행하기 위해 파이썬(Python)의 'statsmodel' 모듈을 사용하였고, 최대가능도방법(Maximum Likelihood Estimation, Song et al., 2010)을 적용하여 확률예측모델을 도출하였다. 그러나 대부분의 데이터는 이러한 기계학습 모델을 바로 도출하기에 적합하지 않다. 따라서 CSMS 데이터를 이용한 기계학습 결과를 소개하기에 앞서 데이터 전처리의 중요성 및 필요성에 대해 언급하고자 한다.

데이터 전처리

앞서 소개된 CSMS 데이터는 수치데이터와 더불어 문자형(범주형) 데이터를 포함하기 때문에, 이를 수치로 치환하는 과정이 필수적이다. 예를 들어, '조치' 항목은 '응급'과 '미응급'으로 나뉘는데 이를 각각 '0'과 '1'로 치환할 수 있다. 또 다른 예로, '주변지형' 항목은 '구릉', '산악', '준산악', '평지'와 같이 네 개의 클래스를 가지고 있으며 이는 각각 '0', '1', '2', '3'과 같이 임의의 숫자로 치환이 가능하다. 다른 범주형 데이터들(KICT, 2019)도 각 클래스 수만큼 0부터 시작하여 '1, 2, 3, ...'과 같이 양의정수로 치환되는 과정을 거친다. 이외에 '최대높이'와 같은 수치데이터는 일반 로지스틱 회귀분석과 같이 치환 혹은 구간 설정 없이 그대로 사용되었다(Lee et al., 2013). 이와 같이 모든 문자형 변수들에 대해 수치로 변환되어야만 기계학습이 가능하다. 이러한 기본적인 치환 과정을 거친 후 기계학습을 수행하면 대부분 데이터 자체 문제로 인해 수행이 중단되는 경우가 많을 것이다. 다음으로 이러한 문제에 대해 소개하고자 한다.

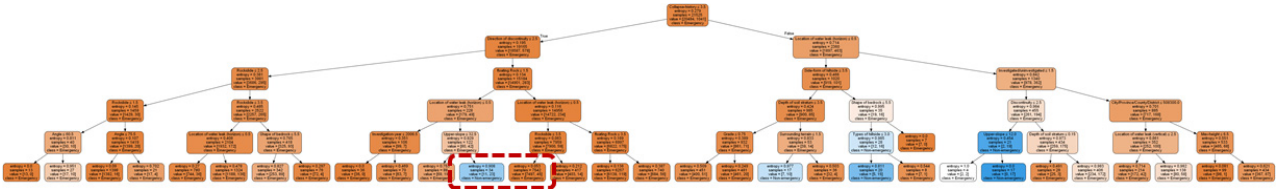
최근 주목받고 있는 빅데이터와 비교한다면 CSMS데이터는 상대적으로 적은 양이지만, 사람이 기록하기 때문에 여러 누락 및 오기입들이 내포되어 있을 수밖에 없고, 이러한 오기입을 찾아내는 것과 누락된 정보를 유추하는 것은 매우 어려운 작업이다. 본 연구에서는 이러한 데이터 문제를 크게 두 가지로 분류하였다. 첫 번째는 누락이며, 두 번째는 오기입 문제이다. 누락문제는 기계학습이 진행되어 확률모델을 구축하고 이를 활용하기 전에는 과학적인 방법으로는 해결이 어렵다. 수동적으로 누락된 정보에 대한 자료를 찾아서 보완하는 방법이 있을 것으로 예상되지만, 본 연구에서는 우선 누락된 데이터는 제외하고 로지스틱 회귀분석을 수행하였다.

다음으로, 오기입 문제는 누락문제와는 달리 극복할 수 있는 문제이지만, 많은 데이터를 일일이 찾아야하는 번거로움이 있다. 오기입 사례로, 1) '절리'가 '졀리'로 기입된 경우, 2) 'wet'이 'weT'으로 기입된 경우, 3) '1'이 '11'로 기입된 경우가 있을 수 있다. 1)번과 같이 한글 오타의 경우 엑셀에서 쉽게 파악이 가능하며 일괄적인 수정이 가능하다. 반면에, 2)번 같은 경우는 엑셀에서 분류가 쉽게 되지 않는다. 2)번 문제는 문자형 데이터를 수치로 변환하는 과정에서, 수치로 변환되지 못하여 NaN(Not a Number)으로 표시되기 때문에 이러한 행을 일일이 찾아 고치는 방식으로 오기입을 수정하였다. 파이썬에서는 이러한 NaN 값을 일률적으로 다른 숫자로 치환하는 방법이 있지만, 이는 해당 사면 정보를 사용하지 않는 것이기 때문에 추천하지 않는다.

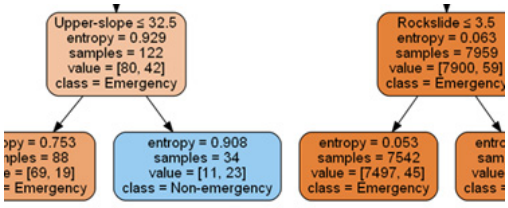
마지막으로 3)번과 같은 경우 또한 잘 드러나지 않으며, 이상함을 발견한 경우에도 누락문제와 마찬가지로 올바른 값을 추정하기가 쉽지 않다. 다행인 점은 이러한 수치 오기입 문제에 해당하는 데이터 수가 그렇지 않은 정상적으로 기입된 데이터 수에 비해 매우 적으며, 구축된 확률모델에 치명적인 영향을 미치지 못 한다는 점이다. 하지만, 이러한 수치적으로 특이함을 보이는 데이터를 사전에 파악하고 관리하는 것은 고품질의 데이터베이스를 구축함에 있어 필요한 부분이다. 이와 같은 특이한 수치 값을 보이는 데이터를 파악하기 위해 비지도학습(Unsupervised Learning)이 사용될 수 있다.

여러 비지도 기계학습 중, 의사결정나무(Decision Tree)는 정해진 의사결정 규칙에 따라 나무구조로 도표하여 분류와 예측을 수행하는 분석 방법이다(Kim et al., 2007). 의사결정나무 분석의 장점은 다차원(여러 항목 혹은 요소) 문제에 대해서도 시각적으로 분류 과정을 쉽게 이해할 수 있다는 점이다. Fig. 3은 29개의 독립변수와 '조치' 항목의 종속변수에 대해 깊이 5의 의사결정나무 분석을 수행한 결과이다. Fig. 3b를 보면 29개의 독립변수에 대한 조치 항목에서 34개의 특이한 데이터 그룹이 있음을 확인할 수 있다.

데이터를 이용한 학습에서는 문자형 데이터의 수치데이터 전환과, 데이터 오기입과 같은 문제 해결이 필수적이다. 하지만 빅데이터와 같이 많은 데이터들이 기록되는 과정에서 인간 오차는 피할 수 없는 부분 중 하나이기 때문에, 완벽히 막는 것보다 그 수를 줄이는 것, 그리고 빠르게 해결하는 절차(Framework)를 제안하는 것이 필요하다. 본 연구에서 제안하는 방안으로는, 1)영문 대문자 혹은 소문자로의 통일, 2) 의사결정나무와 같은 비지도학습을 통한 특이점 데이터 그룹의 파악 및 검토 등이 있다.



(a) Full-diagram



(b) Part for the rare event

Fig. 3. Example of decision-tree analysis.

로지스틱 회귀분석 적용

전체 데이터 기반 '조치' 예측모델 구축

앞선 CSMS데이터 개요 소개에서 언급된 29개의 사면에 대한 정보형 항목을 독립변수로 가정하였고, 판단형 항목인 '조치' 항목을 종속변수로 가정하였다. 교차 검증(Cross-Validation)을 위해 학습데이터와 검증 데이터를 9:1의 임의의 비율로 설정하였다. 이때, 검증 데이터의 비율을 낮게 설정할수록 정확도는 높아질 수 있으나, 데이터 과적합(Over-fitting) 문제가 발생할 수 있다. 로지스틱 회귀분석 수행 결과, 정확도 95.29%의 로지스틱 확률모델이 구축되었고, Table 1에 상세 결과 값을 표현하였다.

Table 1은 29개의 독립변수의 계수와 각 계수들의 신뢰도 구간, 유의확률(p-value)을 나타내며, 유의확률 값의 의미는 아무런 관련이 없음에도 유의미하게 나올 확률 값으로 이해할 수 있다. Table 1을 보면, 유의확률 값이 0.05보다 큰 변수들을 굵은체로 표시하였다. '왕복/편도', '차선수', '각도', '누수위치 가로', '사면형상', '측면형상', '계곡부', '암종' 8개 변수의 유의확률이 상당히 크게 나왔고, 이는 '조치' 종속변수를 추정함에 있어 합리적인 변수가 아님을 의미한다. 따라서 이 8개의 변수를 제외하고 다시 로지스틱 회귀분석을 수행하였다.

Table 1. The prediction model for 'Action' (using 29 independent variables)

	Coef.	Confidence interval		z	$P>\|z\|$
		2.5%	97.5%		
Constant	-3.8290	-3.921	-3.737	-81.432	0.000
Province/City/Town/District	-0.0732	-0.144	-0.002	-2.029	0.042
Round/One-way	**0.0147**	**-0.050**	**0.079**	**0.448**	**0.654**
No. of way	**-0.0762**	**-0.161**	**0.008**	**-1.768**	**0.077**
Date	-0.1290	-0.214	-0.044	-2.977	0.003
Survey/Non-survey	-0.3116	-0.365	-0.258	-11.44	0.000
Slope length	0.1055	0.040	0.171	3.152	0.002
Max. height	0.1428	0.074	0.212	4.074	0.000
Angle	**0.0734**	**-0.020**	**0.166**	**1.545**	**0.122**
Gradient	-0.3170	-0.488	-0.146	-3.640	0.000
Gradient of upper-slope	0.1235	0.070	0.177	4.560	0.000
Distance for rockfall	-0.3674	-0.454	-0.281	-8.341	0.000
Berm	-0.1565	-0.243	-0.070	-3.557	0.000
Slope material	-0.2069	-0.285	-0.129	-5.199	0.000
Topography	-0.0821	-0.152	-0.013	-2.315	0.021
Ground water	0.1473	0.093	0.202	5.320	0.000
Loc. of leaking (top-bottom)	0.3327	0.236	0.430	6.710	0.000
Loc. of leaking (left-right)	**0.0595**	**-0.035**	**0.155**	**1.229**	**0.219**
Weathering grade	0.1519	0.062	0.241	3.328	0.001
Discontinuities' type	-0.3712	-0.441	-0.301	-10.405	0.000
Slope-shape	**0.0116**	**-0.044**	**0.068**	**0.406**	**0.684**
Slide-shape	**-0.0160**	**-0.077**	**0.045**	**-0.517**	**0.605**
Valley in slope	**0.0164**	**-0.037**	**0.070**	**0.600**	**0.549**
Collapse record	0.5877	0.537	0.638	22.749	0.000
Floating rock	-0.1125	-0.175	-0.050	-3.524	0.000
Rock fall	0.4388	0.368	0.510	12.124	0.000
Lithology	**-0.0388**	**-0.099**	**0.022**	**-1.257**	**0.209**
Soil depth	0.2055	0.147	0.264	6.910	0.000
Bedrock-shape	-0.1985	-0.256	-0.141	-6.818	0.000
Type of discontinuities	-0.0944	-0.173	-0.016	-2.368	0.018

Table 2는 로지스틱 회귀분석을 재 수행한 결과이다. 결과 값들을 살펴보면, 유의확률이 큰 독립변수는 더 이상 나타나지 않았다. 하지만, 정확도는 95.29%에서 95.48%로 크게 개선되지는 않았다. 제외된 8개의 독립변수들의 계수 자체가 작은 값을 가졌었고, 이는 종속변수인 '조치' 항목과 큰 연관성이 없음을 의미한다. 이와 달리, '조치' 변수와 관련이 클 것으로 예상되는 '붕괴이력', '낙석'과 같은 변수들을 보면, 상대적으로 큰 계수 값을 갖는 것을 확인할 수 있다. 따라서 유의확률이 큰 변수들을 제거하는 것은 타당하지만 애초에 관련성이 적은 변수들이 제외되었기 때문에, 정확도가 크게 개선되지 않은 것으로 이해할 수 있다. 그럼에도 불구하고 약 95%의 수치는 매우 높은 정확도이기 때문에, 독립변수에 대한 사전정보가 있다면 높은 신뢰도를 갖는 종속변수 예측이 가능하다.

Table 2. The prediction model for 'Action' (using 21 independent variables)

	Coef.	Confidence interval		z	$P>\|z\|$
		2.5%	97.5%		
Constant	-3.8021	-3.893	-3.711	-82.237	0.000
Province/City/Town/District	-0.1002	-0.171	-0.030	-2.788	0.005
Date	-0.0905	-0.170	-0.011	-2.219	0.026
Survey/Non-survey	-0.3242	-0.377	-0.272	-12.143	0.000
Slope length	0.0898	0.028	0.152	2.841	0.004
Max. height	0.1399	0.071	0.209	3.964	0.000
Gradient	-0.4316	-0.549	-0.314	-7.178	0.000
Gradient of upper-slope	0.1340	0.078	0.190	4.652	0.000
Distance for rockfall	-0.4103	-0.507	-0.314	-8.352	0.000
Berm	-0.174	-0.257	-0.091	-4.128	0.000
Slope material	-0.2251	-0.302	-0.148	-5.712	0.000
Topography	-0.1083	-0.176	-0.040	-3.122	0.002
Ground water	0.1460	0.092	0.200	5.306	0.000
Loc. of leaking (top-bottom)	0.3601	0.288	0.433	9.721	0.000
Weathering grade	0.1545	0.066	0.243	3.418	0.001
Discontinuities' type	-0.3646	-0.433	-0.296	-10.405	0.000
Collapse record	0.5994	0.549	0.650	23.323	0.000
Floating rock	-0.1058	-0.169	-0.043	-3.309	0.001
Rock fall	0.4477	0.377	0.519	12.375	0.000
Soil depth	0.1897	0.130	0.249	6.262	0.000
Bedrock-shape	-0.1977	-0.255	-0.141	-6.779	0.000
Type of discontinuities	-0.0812	-0.158	-0.004	-2.073	0.038

차년도 '조치' 예측모델 구축

다음으로는 차년도 '조치' 항목 변수를 예측하기 위한 로지스틱 모델을 구축하였다. 이전 회귀분석에서 구축된 확률모델은 2006년부터 2019년까지의 모든 데이터를 사용하여, 로지스틱 회귀분석의 정확도를 확인하였다. 이번에 소개하는 회귀분석에서는 이전 몇 개년도의 데이터를 학습데이터로, 바로 다음 1개년도의 데이터를 검증 데이터로 가정함으로써, 이전 데이터들을 활용한 향후 '조치' 항목 예측 가능성을 검토하기 위해 수행되었다.

총 13개의 경우를 가정하고 로지스틱 회귀분석을 수행하였다. Table 3을 보면, 2006년 데이터를 학습데이터로 사용하여 로지스틱 모델을 구축하고, 구축된 모델에 대해 2007년 데이터를 이용하여 모델의 정확도를 계산하였다. 이와 같이, 마찬가지로 2006년부터 특정 년도까지의 데이터들을 학습데이터로 가정하여 로지스틱 모델을 구축한 후, 차년도 데이터를 이용하여 구축모델을 검증하였다.

Case 6의 경우 2006년부터 2011년까지의 데이터를 학습데이터로 가정하였고, 이를 사용하여 구축된 모델에 대해 2012년 데이터를 사용하여 정확도를 계산한 결과 약 88%로 다른 경우들에 비해 낮음을 확인하였다. 이는 앞서 조치 예측모델보다도 낮은 정확도이다. 이 정확도가 의미하는 바는, 2012년도에 새로 수집된 사면의 '조치' 항목의 응급·미응급 비율 양상이 이전과는 상이함을 의미한다. 이는 다시 크게 두 가지 관점으로 볼 수 있다. 첫 번째는 실제로 이전 비탈면 조치

결과들과는 달리, 새로운 지역 혹은 특이점이 있어서 실제로 응급·미응급 비율이 다른 경우이고, 두 번째는 전문가의 판단에 실수가 있는 경우이다. 어떠한 경우라도 상관없이, 이전 데이터들과의 상이함을 파악하고 이를 통해 새로 수집된 데이터 검토가 가능하다.

Table 3. Demonstration of logistic models to predict the next year data

	Learning-data's date	No. of learning-data	Validation-data's date	No. of validation-data	Accuracy (%)
Case 1	2006	4938	2007	5625	94.33
Case 2	2006~2007	10563	2008	7025	94.32
Case 3	2006~2008	17588	2009	6013	97.16
Case 4	2006~2009	23601	2010	3161	97.41
Case 5	2006~2010	26762	2011	820	96.22
Case 6	**2006~2011**	**27582**	**2012**	**531**	**88.14**
Case 7	2006~2012	28113	2013	470	97.23
Case 8	2006~2013	28583	2014	305	97.70
Case 9	**2006~2014**	**28888**	**2015**	**353**	**99.43**
Case 10	2006~2015	29241	2016	627	98.72
Case 11	2006~2016	29868	2017	179	96.65
Case 12	**2006~2017**	**30047**	**2018**	**258**	**99.22**
Case 13	2006~2018	30305	2019	446	97.76

반면에, Case 6과는 달리 Case 9 와 12의 경우는 매우 높은 정확도를 보여주었다. 이는 각 경우의 해당년도에 수집된 데이터가 이전 년도를 기반으로 구축된 모델과 매우 유사함을 의미한다. 다음으로는, 이전 데이터들과 상이함을 보여준 2012년 이후 데이터들을 이용하여 예측 검증을 수행하였다. Table 4는 3개년도 데이터를 학습데이터로 가정한 경우들에 대한 결과이며, Table 5는 2개년도 데이터를 학습데이터로 사용한 결과이다.

Table 4. Demonstration of logistic models using the data for the previous three year

	Learning-data's date	No. of learning-data	Validation-data's date	No. of validation-data	Accuracy (%)
Case 1	2013~2015	1128	2016	627	98.56
Case 2	2014~2016	1285	2017	179	96.65
Case 3	2015~2017	1159	2018	258	98.84
Case 4	2016~2018	1064	2019	446	97.76

위 결과들을 살펴보면, 2개년도 데이터들을 이용하여 학습시킨 로지스틱 모델이 3개년도 데이터를 이용한 경우보다 대체로 정확도가 높은 것을 확인할 수 있다. 이는 예측하고자 하는 년도와 가까운 년도의 데이터가 양상이 비슷하며 사용하기에 적합함을 의미한다. 또한, 사용된 학습 및 검증 데이터의 수를 보면, 상당히 적음에도 불구하고 구축된 모델의 정확도가 높다. 이는 데이터 자체가 일관성이 있음을 시사한다. 다음 식은 Table 5의 Case 1에 대해 도출된 로지스틱 회귀분석 수식이다. 여기서 변수 x_1, x_2, ..., x_{29}는 앞선 CSMS데이터 개요 소개에서 언급한 29개의 독립변수(순서대로 '시/도/군/구', '왕복/편도', ..., '불연속면')를 의미한다.

Table 5. Demonstration of logistic models using the data for the previous two year

	Learning-data's date	No. of learning-data	Validation-data's date	No. of validation-data	Accuracy (%)
Case 1	**2013~2014**	**775**	**2015**	**353**	**99.43**
Case 2	2014~2015	658	2016	627	98.56
Case 3	2015~2016	980	2017	179	96.09
Case 4	**2016~2017**	**806**	**2018**	**258**	**99.22**
Case 5	2017~2018	437	2019	446	97.31

$$\begin{aligned}\mathrm{logit}(p) = & -2.9873 - 0.0443x_1 - 0.2365x_2 - 0.4577x_3 + 0.1576x_5 + 0.3077x_6 - 0.1058x_7 - 0.0933x_8 \\ & + 0.0506x_9 - 0.1475x_{10} - 0.4459x_{11} + 0.3851x_{12} + 0.1215x_{13} - 0.0897x_{14} - 0.1377x_{15} \\ & + 0.1388x_{16} - 0.1614x_{17} + 0.0722x_{18} - 0.1195x_{19} - 0.0437x_{20} - 0.1764x_{21} - 0.2316x_{22} \\ & - 0.0708x_{23} + 0.0489x_{24} - 0.0026x_{25} - 0.3533x_{26} - 0.1978x_{27} + 0.1915x_{28} - 0.3346x_{29}\end{aligned} \quad (4)$$

'계측추천' 예측모델 구축

'조치' 항목이외에 '계측추천' 항목 또한 두 클래스로 이루어진 판단형 변수이다. 이에 대해서도 조치 예측모델 구축 시와 마찬가지로 2006년부터 2019년까지의 모든 데이터를 이용하여 로지스틱 모델을 구축해보았다. 동일한 데이터 전처리 과정을 수행하였고, 마찬가지로 9:1의 교차검정 비율을 설정하였다. Table 6은 구축된 예측모델의 상세결과이며, 정확도는 99.54%로 매우 높은 정확도를 보였다. 이 예측 모델의 경우 매우 높은 정확도 덕분에, '조치' 예측모델과 달리 전체년도 데이터를 이용하여 구축된 모델을 차년도 예측모델로 활용 가능할 것으로 생각된다. Table 7은 유의확률이 높은 변수들을 제외한 결과이다.

Table 6. The prediction model for 'Measuring-Recommendation' (using 29 independent variables)

	Coef.	Confidence interval		z	$P>\|z\|$
		2.5%	97.5%		
Constant	-6.4806	-6.797	-6.165	-40.188	0.000
Province/City/Town/District	**0.0785**	**-0.119**	**0.276**	**0.779**	**0.436**
Round/One-way	**-0.0125**	**-0.250**	**0.225**	**-0.103**	**0.918**
No. of way	0.2720	0.051	0.493	2.410	0.016
Date	-0.5377	-0.782	-0.293	-4.308	0.000
Survey/Non-survey	-0.1731	-0.348	0.002	-1.938	0.053
Slope length	0.1516	0.039	0.264	2.650	0.008
Max. height	0.4008	0.314	0.488	9.052	0.000
Angle	**-0.2618**	**-0.966**	**0.443**	**-0.728**	**0.467**
Gradient	**-0.6731**	**-1.693**	**0.347**	**-1.294**	**0.196**
Gradient of upper-slope	0.1050	0.010	0.200	2.165	0.03
Distance for rockfall	0.0891	0.036	0.142	3.281	0.001
Berm	0.3312	0.208	0.454	5.265	0.000
Slope material	**0.0320**	**-0.198**	**0.262**	**0.274**	**0.784**
Topography	**0.0376**	**-0.172**	**0.247**	**0.351**	**0.725**
Ground water	0.1651	-0.001	0.331	1.951	0.051

Table 6. Continued

	Coef.	Confidence interval		z	$P>\|z\|$
		2.5%	97.5%		
Loc. of leaking (top-bottom)	**0.0948**	**-0.235**	**0.425**	**0.564**	**0.573**
Loc. of leaking (left-right)	**0.0530**	**-0.278**	**0.384**	**0.314**	**0.753**
Weathering grade	0.3564	0.091	0.622	2.635	0.008
Discontinuities' type	-0.4685	-0.679	-0.258	-4.363	0.000
Slope-shape	**-0.1017**	**-0.274**	**0.071**	**-1.154**	**0.248**
Slide-shape	-0.3994	-0.598	-0.200	-3.934	0.000
Valley in slope	-0.1411	-0.305	0.022	-1.692	0.091
Collapse record	0.2716	0.113	0.430	3.355	0.001
Floating rock	-0.3505	-0.518	-0.183	-4.103	0.000
Rock fall	**-0.0415**	**-0.225**	**0.142**	**-0.443**	**0.658**
Lithology	**0.0029**	**-0.173**	**0.179**	**0.033**	**0.974**
Soil depth	-0.5318	-0.893	-0.171	-2.888	0.004
Bedrock-shape	**0.1047**	**-0.060**	**0.270**	**1.245**	**0.213**
Type of discontinuities	0.3756	0.107	0.645	2.736	0.006

Table 7. The prediction model for 'Measuring-Recommendation' (using 17 independent variables)

	Coef.	Confidence interval		z	$P>\|z\|$
		2.5%	97.5%		
Constant	-6.3351	-6.624	-6.046	-42.995	0.000
No. of way	0.3006	0.089	0.512	2.791	0.005
Date	-0.5666	-0.798	-0.335	-4.798	0.000
Survey/Non-survey	-0.2058	-0.376	-0.035	-2.364	0.018
Slope length	0.1807	0.065	0.297	3.055	0.002
Max. height	0.3601	0.279	0.441	8.698	0.000
Gradient of upper-slope	0.2108	0.033	0.388	2.329	0.020
Distance for rockfall	0.1754	0.086	0.265	3.856	0.000
Berm	0.3421	0.227	0.457	5.816	0.000
Ground water	0.1748	0.041	0.308	2.568	0.010
Weathering grade	0.3693	0.126	0.613	2.971	0.003
Discontinuities' type	-0.4163	-0.618	-0.215	-4.047	0.000
Slide-shape	-0.4343	-0.622	-0.246	-4.526	0.000
Valley in slope	-0.1624	-0.326	0.001	-1.948	0.051
Collapse record	0.2044	0.050	0.359	2.594	0.009
Floating rock	-0.3414	-0.490	-0.192	-4.491	0.000
Soil depth	-0.4371	-0.717	-0.157	-3.055	0.002
Type of discontinuities	0.3616	0.111	0.612	2.831	0.005

결 론

본 연구는 도로비탈면관리시스템이 관리하는 기초조사 및 정밀조사 자료를 바탕으로 기계학습인 로지스틱 회귀분석을 통해 가지고 있는 데이터 활용 가능성을 검토하고자 수행되었다. 기계학습을 수행하기 위해 데이터 전처리를 수행한 후, 독립변수와 종속변수를 선정하여 로지스틱 모델을 도출하였다. 도출된 모델들은 매우 높은 정확도를 보였는데, 이는 모델 구축 시 사용된 변수들에 해당하는 데이터들이 일관성 있게 기록되었음을 의미한다. 다음으로 향후 기초·정밀 조사를 수행해야하는 신규 사면이 발생한 경우, 일부 정보를 이용하여 응급·미응급을 판단하는 조치 항목을 예측 가능한지 검토하고자 하였다. 검토 결과, 구축된 모델은 매우 높은 정확도를 보였으며 사전 예측뿐만 아니라 실제 정밀조사 정보를 기록할 때, 구축된 모델에서 계산된 예측값과 비교함으로써 상이함이 없는지 검토 가능하다. 즉, 필터링 기능을 수행할 수 있으며 이를 통해 고품질의 데이터베이스 구축이 가능할 것으로 기대된다.

CSMS에서 관리하고 있는 전국 도로 비탈면에 대한 기초·정밀 조사 데이터는 많은 사면 현황과 향후 유지관리 의사결정을 위한 정보들을 가지고 있다. 그런데, 인간오차로 인해 누락된 정보와 혹은 오기입된 정보로 인해 기계학습 적용이 어렵고, 데이터가 가지고 있는 숨겨진 정보들을 발굴하고 활용하기란 쉽지 않다. 본 연구에서 수행한 로지스틱 회귀분석을 통해, CSMS 데이터 활용 가능성은 검토하였지만, 사용된 독립변수 및 종속변수 외에 누락된 정보들이 많은 항목(변수)들은 활용되지 못했다. 따라서 추후 누락된 데이터를 보완하고 활용하기 위한 후속 연구가 필요할 것으로 생각된다.

사 사

본 연구는 국토교통부와 국토교통과학기술진흥원의 지원(과제번호: 20SCIP-C146569-03)을 받아 수행되었습니다. 이에 감사드립니다.

References

Baek, S.A., Cho, K.H., Hwang, J.S., Jung, D.H., Park, J.W., Choi, B., Cha, D.S., 2016, Assessment of slope failures potential in forest roads using a logistic regression model, Journal of Korean Forest Society, 105(4), 429-434 (in Korean with English abstract).

Chae, B.G., Kim, W.Y., Cho, Y.C., Kim, K.S., Lee, C.O., Choi, Y.S., 2004, Development of a logistic regression model for probabilistic prediction of debris flow, The Journal of Engineering Geology, 14(2), 211-222 (in Korean with English abstract).

KICT (Korea Institute of Civil Engineering and Building Technology), 2019, Operation of road cut slope management system in 2018, Republic of Korea Ministry of Land, Infrastructure and Transport, 355p.

Kim, T.H., Shin, Y.S., Lee, U.K., Kang, K.I., 2007, Decision support model using a decision tree for framework selection in tall building construction, Journal of the Architectural Institute of Korea Structure & Construction, 23(11), 177-184 (in Korean with English abstract).

Lee, H., Kim, G., 2012, Landslide risk assessment in Inje using logistic regression model, Korea Society of Surveying, Geodesy, Photogrammetry, and Cartography, 30(3), 313-321 (in Korean with English abstract).

Lee, J.H., Chang, B.S., Kim, Y.S., Suk, J.W., Moon, J.S., 2013, Risk assessment for large-scale slopes using multiple regression analysis, Journal of the Korean Geotechnical Society, 29(11), 99-106 (in Korean with English abstract).

Lee, Y.H., Kim, J.R., 2004, A proposal of the evaluation method for rock slope stability using logistic regression analysis, Korean Society for Rock Mechanics, 14(2), 133-144 (in Korean with English abstract).

Quan, H.C., Lee, B.G., Lee, C.S., Ko, J.W., 2011, The landslide probability analysis using logistic regression analysis and artificial neural network methods in Jeju, Journal of Korean Society for Geospatial Information Science, 19(3), 33-40 (in Korean with English abstract).

Song, J., Kang, W.H., Kim, K.S., Jung, S., 2010, Probabilistic shear strength models for reinforced concrete beams without shear reinforcement, Structural Engineering & Mechanics, 34(1), 15-38.

Yeon, Y.K., 2011, Evaluation and analysis of Gwangwon-do landslide susceptibility using logistic regression, The Korean Association of Geographic Information Studies, 14(4), 116-127 (in Korean with English abstract).