

머신러닝 기반 시설물 안전 점검·진단용역 부실 판정 요인에 대한 연구

Investigating Factors Contributing to Inadequate Facility Safety Inspections and Diagnosis Services: A Machine Learning Approach

박준용¹, 송지훈^{2*}

Junyong Park¹, Chie Hoon Song^{2*}

〈Abstract〉

Evaluating the adequacy of facility safety inspection and diagnosis services performed by private enterprises is a time-consuming and administratively complex process. This study aims to analyze the determinants that could influence the rating of these safety inspection and diagnosis services using data analytics approach. Through a comparative analysis of several machine learning algorithms suitable for multi-class classification, we selected the model with the best performance (Random Forest) and identified the main determinants using the permutation importance technique. Among the variables examined, "contract value," "days of service performed" and "adherence to fair market value" were found to be strongly correlated with the rating assessments. Furthermore, we discovered that the skills and expertise of service performing personnel significantly impacted the rating. The results of this study can contribute to the enhancement of the current post-evaluation administrative processes and offer valuable insights into rating assessments by incorporating previously unexplored variables pertaining to both service providers and the services itself.

Keywords : Machine Learning, Prediction Model, Big Data, Permutation Importance, Hyper-Parameter Tunning, Facility Safety

1 제1저자, 박사과정, 경상국립대학교 대학원 기술경영학과
E-mail: yuchicop@gnu.ac.kr

2* 교신저자, 조교수, 경상국립대학교 대학원 기술경영학과
E-mail: chsong01@gnu.ac.kr

1 First author, Graduate Student (Ph.D. program), Gyeongsang National University, Department of Management of Technology

2* Corresponding author, Assistant Professor, Gyeongsang National University, Department of Management of Technology

1. 서론

최근 처벌보다는 예방이 우선시 되는 안전 및 안심 사회 구축을 위한 다양한 제도적 논의의 중요성이 대두되고 있다. 특히, 지속 가능하고 안전한 사회의 조성은 정부에서 중점을 두고 있는 중요한 정책 집행 방향이기도 하다. 그러나 대한민국은 경제 수준에 비해 높은 안전불감증과 낮은 안전의식을 나타내고 있다. 통계청이 발표한 『근로자 10만명당 치명적 산업재해 수(OECD회원국)』에 의하면 2021년을 기준으로 한국은 공동 3위를 기록했으며 [1], 2022년에도 산업재해로 인한 사망자 수가 건설업의 경우 341명이 발생하였고 이는 전체 산업재해 사망자 수의 53%에 달하였다[2]. 이처럼 안전 리스크에 대한 인식 부족은 사회 전반에서 널리 퍼져 있는 현상으로 볼 수 있다. 현 정부는 국정과제 69번의 수행을 통해 “국민이 안심하는 생활안전 확보”를 추진하고 있다. 주요 내용으로는 건설 주체(발주·시공·감리)의 안전 확보 책무 강화, 건설 현장 스마트 안전장비 확대, 시설물 안전관리 강화 등을 통해 국민이 안심할 수 있는 생활환경 조성을 위한 국가적 대책을 마련하고자 한다.

다중이 이용하는 기반시설물인 교량, 터널, 건축물 등은 정기적으로 정밀점검과 상위 단계인 정밀안전진단을 의무적으로 수행해야 한다. 이는 시설물의 안전성을 정기적으로 확인하고 안전에 영향을 주는 결함을 찾아 적시에 필요한 보수와 보강을 시행하기 위함이다. 결함이 장기간 방치되거나 필요한 보수 또는 보강이 누락되는 문제를 예방하고자 시설물 관리주체는 시설물의 안전과 기능을 유지하기 위하여 정기적으로 안전점검·진단(이하 점검·진단)을 실시해야 하며, 안전진단 전문기관 및 유지관리업자가 이를 대행할 수 있다. 하지만 이러한 점검을 대행하는 업체들이 “복제·거짓·부실 작성의 판단기준”을 위반하여 결과보고서를 작

성하는 경우가 적발되었고, 이는 부실시공만큼 시설물의 안전을 위협하는 요인으로 지적되었다. 예를 들어, 보고서 제출에 있어 과거 보고서 사진을 그대로 재사용하는 등 부실한 보고서 작성 관행이 근절되지 않고 있다[3]. “정밀점검 및 정밀안전진단 실시 결과에 관한 평가제도(이하 평가제도)”는 이러한 점검·진단의 부적정성을 개선하기 위해 2002년 처음 도입되었다. 점검·진단 부실의 문제점은 단순히 점검 및 진단에 드는 비용이 낭비되는 것을 떠나 이를 바탕으로 부적절한 보수·보강이 시행되는 경우 추가적인 손실 비용이 발생하기 때문이다[4]. 그 외에도 부실한 점검·진단으로 적시에 보수·보강이 이뤄지지 않는다면, 시설물의 성능과 수명에도 부정적 영향을 미칠 수 있다.

현 평가제도는 시설물 점검·진단 완료 후 제출된 보고서에 대한 적정성을 검토하고 보완을 요청하는 사후 방식이다. 사후평가 대상으로는 이전 점검에서 2단계 이상 등급 변동이 있는 경우 선정한다[5]. 이후 평가에서 총점이 60점 미만이거나 중요 평가항목이 과락인 경우 부적정 처리된다. 이 경우 법령에 따라 정해진 기간 내에 보고서를 보완하여 다시 제출하여야 한다[6]. 현재 평가방식은 추가 예산, 행정력, 인력 등의 자원이 소모되는 복잡한 절차를 수반한다. 또한, 안전점검이 보완 될 때까지 부적정 판정을 받은 시설물에 의해 발생하는 안전 위협이 존재한다.

이에 반하여 정형화된 데이터 기반 예측 모델을 통해 민간이 수행한 시설물 안전 점검·진단 과업의 부실 또는 적정 수행 등급을 부여한다면, 보고서가 제출되는 즉시 데이터가 확보됨에 따라 더 신속한 평가가 가능하다. 즉, 부실한 과업 수행이 예상되는 시설물 점검·진단이 제출되었을 때 사후평가라는 행정 절차를 우회하여 집중 관리를 위한 예산과 인력의 우선 투입이 가능하다. 이러한 예측 모델을 효과적으로 활용하기 위해서는 제출된

용역보고서의 등급 판정에 영향을 미치는 요인에 대한 분석이 필요하다. 본 연구에서는 부실 가능성이 있는 용역 등급평가 결과와 관련 데이터에 대한 검토를 기반으로 머신러닝 기반 예측 모델을 구축하고, 이를 기반으로 예측성능에 가장 중대한 영향을 미치는 요인을 파악하고자 한다. 이를 통해 향후 부실한 점검·진단 예방에 기여 가능한 변수를 식별하고 궁극적으로 시설물 관련 사고 예방 및 안전 향상에 기여할 수 있는 전략안을 제시하고자 한다. 예측 모델 구현에 있어 시설물 관련 변수만으로는 시설물 점검·진단에 대한 등급 예측이 어려울 것으로 판단되어, 시설물 관련 변수 외 용역업체와 용역 자체와 관련 변수들을 추가로 고려하고자 한다. 본 연구는 완료된 점검·진단 용역의 등급 평가를 기반으로 한 요인 분석을 목표로 하기에, 다중 분류에 특화된 머신러닝 모델을 구축해야 하며, 적합한 알고리즘들의 비교를 통해 최적의 모델을 선정하고자 한다. 특히, 기존 연구에서 다루지 않았던 변수들을 포함해 모델을 생성하고 현 평가방식의 단점을 보완하는 데 기여할 수 있다는 점에서 학술적 및 실무적 의의를 지닌다.

본 논문의 구성은 다음과 같다. 2장에서는 시설물 안전 등급 평가 및 안전 위험요인 발굴 관련 선행연구를 검토한다. 3장에서는 분석에 활용한 데이터에 대한 설명을 제공하고 단계별 분석 프로세스에 대해 기술한다. 4장에서는 논문의 분석 결과를 예측 모델의 정확성 및 결정요인 측면에서 제시하고 논의한다. 마지막 장에서는 연구의 결론과 한계점 및 후속 연구의 방향성에 대해 논의한다.

2. 시설물 안전 등급 평가 및 안전 위험요인 발굴 관련 선행연구

시설물 안전 등급 평가 및 안전 위험요인 발굴

관련 선행연구는 (1) 설문지 기반 실증분석, (2) 텍스트 마이닝, (3) 빅데이터 분석 등으로 구분할 수 있다. 그중 설문지 기반 선행연구는 다음과 같다. 한범진 외(2023)는 건설산업에서 지속가능한 재해예방을 위해 현장의 안전 분위기 향상을 위한 노력이 필요함을 강조하였다. 이를 검증하기 위해 건설 현장의 안전 전문가를 대상으로 AHP 설문조사를 실시해 현장 안전 분위기 평가 요인의 상대적 중요도를 분석하였다. 그 결과, ‘현장관리자의 안전 중시 수준’, ‘현장참여자들 간의 안전에 대한 신뢰 수준’, ‘안전에 대한 의사소통 수준’ 등이 상대적으로 높은 중요도를 보였으며, 도출된 우선순위를 기준으로 건설 현장의 안전 분위기 향상을 위한 개선 방안을 제시하였다[7]. 정인수(2014)는 소규모 시설물에 대한 안전관리 강화를 목적으로 전문가 자문, 설문조사 및 사고사례분석을 통해 시설물별 안전관리 강화 범위를 논하였다. 소규모 시설물 중 교량, 터널, 항만, 건축물, 하천, 옹벽 및 절토사면 등이 분석 대상으로 선정되었다. 공중에 대한 위해 가능성이 높은 시설물은 1·2종 시설물로 통합하여 관리할 것을 제안하였다[8].

텍스트 마이닝 활용 선행연구 사례를 살펴보면, 이정민·김율성(2023)은 항만안전 관련 뉴스 기사와 학술지 초록에 대한 토픽모델링 분석을 통해 항만 안전 위험요인을 도출하였다. 식별된 주요 위험요인으로는 보안적, 자연적, 인적, 환경적 및 기계적 요인 등이 있으며, 이들에 대한 지속적 개선을 통해 안전문화를 발전시킬 필요성이 존재함을 시사하였다[9]. 김진국 외(2022)는 터널 교통안전 관련 국내 논문 데이터를 기반으로 키워드 빈도 분석과 토픽 모델링 분석을 수행하였다. 총 10개의 토픽이 도출되었으며, 그중 “과속”, “유고”, “측위”, “고령운전자”, “운전자 심리”와 같은 주제들이 상대적으로 높은 비중을 차지하는 위험요인으로 나타났다[10]. 이와 유사한 맥락에서 이상규(2018)는 토픽모델링을 통해 건설안전사고 관련 뉴스데이터를 분석하였고,

토픽별 시계열 분석을 기반으로 증가추세와 감소추세를 나타내는 안전사고 이슈를 도출하였다[11].

빅데이터 분석을 활용한 선행연구는 다음과 같다. 장태연 외(2023)는 다양한 시설물 유지관리 데이터와 머신러닝 알고리즘을 활용해 시설물 부재상태등급 예측 모델을 구축하였다. 검토된 여러 알고리즘 중 XGBoost 기반 모델이 가장 우수한 예측성능을 나타냈으며, 예측 결과는 시설물 유지관리 의사결정에 요구되는 부재상태 예측정보를 효과적으로 제공할 수 있을 것으로 판단된다[12]. 홍지수·전세진(2023)은 준공 30년 이상인 노후 교량에 대하여 의사결정나무 및 랜덤포레스트 기반 분류 모델을 통해 교량의 안전등급을 예측하는 방법을 제안하였다. 데이터 불균형 문제해결을 위해 다양한 샘플링 기법을 적용한 결과, 노후화된 교량의 안전등급 예측에는 랜덤 언더샘플링 방식의 재현율이 더 우수함을 확인하였다. 또한, 모델에 영향을 미치는 변수 중요도를 파악한 ‘공용기간’, ‘교량연장’, ‘교통량’, ‘시설물종별등급구분’ 등이 주요 요인을 도출되었다[13]. 김정집 외(2023)는 인공지능 기반 사물인터넷 구조물 안전관리 예측 시스템을 개발하였으며, 이 과정에서 LSTM(long short-term memory) 모델을 활용해 높은 수준의 미래 구조물 불안정 상태의 예측이 가능함을 시사하였다[14].

3. 연구 방법

3.1 연구 데이터

본 연구에서는 머신러닝 기반 점검·진단 용역 평가 등급 예측 모델을 구축하기 위해 관련 원본 데이터를 국토안전관리원이 관리 중인 시스템 DB에서 수집하였다. 먼저 점검·진단용역 적정/부실 여부 사후평가를 관리하는 점검진단실시결과 평가

Table 1. Overview of variables

변수 카테고리	변수명	데이터 유형
타겟 변수	용역 평가등급	범주형
시설물 관련 요인	시설물종별	범주형
	시설물중분류	범주형
	시설물담당부처	범주형
	시설물 최종안전등급	범주형
	공공민간관리 주체여부	범주형
용역업체 관련 요인	점검진단기관 유형	범주형
	용역면허등록갯수	수치형
	용역업체소재지	범주형
	고급특급기술인 보유	수치형
	기사 및 기술사 보유	수치형
	석사 및 박사 보유	수치형
	시정명령 및 과태료 부과건수	수치형
영업정지 및 등록취소 건수	수치형	
용역 관련 요인	점검진단 시작일	날짜형
	점검진단 종료일	날짜형
	용역계약금액_천원_log	수치형
	적정대가_천원_log	수치형

시스템(이하 IREMS시스템)을 통하여 사후평가 대상 점검·진단용역 데이터를 수집하였다. 시설물 및 용역업체 관련 데이터는 시설물통합정보관리시스템(이하 FMS시스템)을 통해 수집하였다. 이후 데이터 통합 과정을 거쳐 2018년에서 2023년 사이 실시된 점검·진단 관련 용역에 대한 총 14,891건의 데이터를 확보하였다. 수집한 원본 데이터의 입력 변수명과 데이터 유형은 총 3개의 카테고리로 구분하여 Table 1에 정리하였다. 용역 평가등급은 예측하고자 하는 타겟 변수로 “적정”, “보완”, “부실”, “불량”의 4가지 클래스로 구성된다.

3.2 연구 프레임워크

본 연구는 Fig. 1에 명시된 데이터 분석 프로세스를(총 5단계) 따라 진행되었다. 먼저 앞서 언급

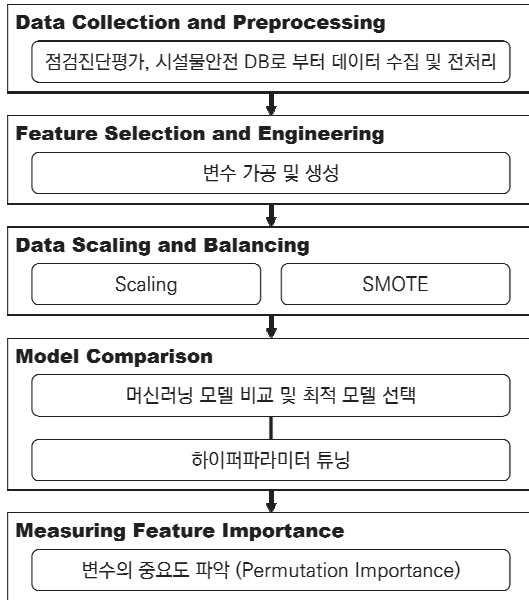


Fig. 1 Outline of the analysis framework

한 시스템 DB를 통하여 데이터를 수집한 후, 이에 대한 전처리를 수행하였다. 두 번째 단계에서는 수집한 데이터에 대한 변수 가공과 파생 변수 생성 작업을 수행하여 모델링에 적합한 형태로 데이터를 정제하였다. 세 번째 단계에서는 데이터에 대한 정규화와 클래스 간 불균형 문제 해소를 위한 균형화(balancing)를 진행하였다. 균형화는 클래스 간 불균형으로 인한 모델의 편향을 방지하고자 수행하였다. 네 번째 단계에서는 “XGBoost”, “RandomForest” 등 다중 분류에 적합한 머신러닝 알고리즘 5개를 기반으로 모델을 생성하고, 각 모델의 예측성능을 비교하였다. 이후 선택된 모델에 대한 하이퍼파라미터 튜닝을 수행하여 모델의 예측성능을 극대화하였다. 마지막으로 가장 우수한 예측성능을 보이는 모델을 바탕으로 용역 평가등급에 영향을 미치는 변수들의 중요도를 도출한 후, 그 결과를 기반으로 주요 결정요인을 규명하였다.

3.3 데이터 수집 및 전처리

본 분석 단계에서는 수집된 원본 데이터 중 결측치가 포함된 행의 경우 다른 값으로 대체가 유의미하지 않을 것으로 판단되어 제외하였다. 나아가 “용역계약금액”과 같은 변수의 경우 결측치를 제거했음에도 값의 분포가 극단적인 성향을 띠었기에, 로그 변환(log-transformation)을 통해 데이터의 분포를 정규분포에 근사하게 변환했다. 나아가 이상치를 포함하는 데이터의 경우 IQR(interquartile range) 기법을 활용해 제거하였다[15]. 그 결과 총 10,283건의 데이터를 획득하였다.

3.4 변수 가공 및 생성

변수 가공 단계에서는 One-Hot-Encoding을 활용해 범주형 데이터를 수치형 데이터로 전환하였다. 파생 변수 생성에 있어서는 “적정대가 준수율”과 “용역 수행일수”를 새롭게 정의하였다. Table 2는 두 변수에 대한 산출식과 데이터 유형을 정리해 나타낸다. “적정대가 준수율”의 경우 “적정대가(천원)” 대비 “용역계약금액(천원)”의 비율을 산출한 값을 의미한다. “용역 수행 일수”는 용역업체가 시행한 “점검진단 종료일”과 “점검진단 시작일”의 차이로 정의된다.

Table 2. Derived variables

변수명	산출식	데이터 유형
적정대가 준수율	(용역계약금액(천원)/ 적정대가(천원))*100	수치형
용역 수행 일수	점검진단 종료일- 점검진단 시작일	수치형

3.5 데이터 정규화 및 균형화

본 단계에서는 수치형 데이터에 적용할 스케일링 기법으로 Min Max Scaling(데이터의 범주를 0과 1 사이의 값으로 변환), Robust Scaling(아웃라이어의 영향을 최소화하는 방식으로 중앙값은 0으로 변환하며, 데이터의 범위를 사분위수 범위로 스케일링), Standard Scaling(데이터를 정규분포에 가깝게 변환)을 검토하였다. 이후 데이터 불균형 문제를 해결하기 위해 본 연구는 오버샘플링 기법의 하나인 SMOTE(Synthetic Minority Over-sampling Technique)를 적용하였다. 이후 예측 대상인 “용역 평가등급” 변수의 불균형 문제를 다루기 위해 오버샘플링 기법 중 하나인 SMOTE(Synthetic Minority Over-sampling TEchnique) 알고리즘을 활용하였다. SMOTE는 분류 문제에서 소수 클래스 데이터의 개수를 증가시키는 기법으로, 기존 소수 클래스 데이터에 가상의 데이터를 생성하여 데이터 불균형을 보완하는 방법이다[16]. 소수 클래스 데이터의 k개 최근접 이웃을 찾은 후, 이웃 데이터들과의 차이 벡터를 계산하고 이를 기반으로 새로운 가상 데이터를 생성한다. 생성된 샘플 데이터는 소수 클래스가 가지는 데이터의 성향을 반영하기에 모델이 더욱 일반화되고 안정적으로 데이터를 학습할 수 있게 도와준다.

3.6 모델 간 성능 비교

본 단계에서는 다중 분류에 널리 활용되는 5개의 머신러닝 알고리즘의 비교를 통해 최적의 성능을 보이는 알고리즘을 선택 후, 하이퍼파라미터 튜닝 과정을 거쳐 모델의 성능을 극대화하였다. 일차적인 필터링을 위해 'XGBoost', 'Random Forest', 'LightGBM', 'Support Vector Machine (SVM)',

'AdaBoost' 알고리즘에 정규화 및 균형화된 입력값을 적용 후 도출된 F1-score 값을 비교하였다. 언급된 알고리즘은 선행연구를 통해 다중 분류 문제에서 우수한 성능을 보이는 것으로 알려진 방법들이다[12, 17]. 선택된 최적 알고리즘에 대해 다양한 조건의 하이퍼파라미터 조합을 탐색하는 하이퍼파라미터 튜닝 과정을 거쳤으며, 이는 파이썬 기반의 scikit-learn(버전 1.2.2) 라이브러리를 활용해 구현하였다. 하이퍼파라미터 조합 탐색 방식으로는 Randomized Search와 Grid Search 중 Randomized Search 방식을 활용하였다. 모델 성능의 비교를 위해 다양한 평가지표를 도출할 수 있으며, 그중 정밀도(precision)와 재현율(recall)의 조화평균을 활용해 모델의 성능을 평가하는 F1-score를 기준으로 전반적인 성능을 비교하였다.

3.7 변수의 중요도 파악

마지막 단계에서는 모델의 예측성능에 영향을 미치는 주요 변수들을 파악하기 위해 변수 중요도 분석을 수행하였다. 이를 위해 Permutation Importance 기법을 활용하였다[18]. 이는 특정 변수가 모델의 예측성능에 미치는 영향을 평가하여 그 중요도를 판별하는 방법이다. 구체적으로는 특정 변수의 값을 랜덤하게 섞어 모델의 성능 변화를 측정한다. 만약 해당 변수를 섞었을 때 모델의 성능이 감소한다면, 그 변수가 예측에 중요한 역할을 한다고 판단할 수 있다. 반대로 성능 변화가 미미하다면 그 변수의 중요도는 낮다고 평가한다. 본 연구에서는 이러한 Permutation Importance 기법을 적용하여 투입한 변수 중에서 용역 평가 등급 예측에 가장 큰 영향력을 미치는 주요 변수들을 도출해 용역 평가 등급 메커니즘을 해석하고자 한다.

4. 연구 결과

4.1 기술통계 분석 결과

본 장에서는 분석한 데이터의 주요 변수에 대

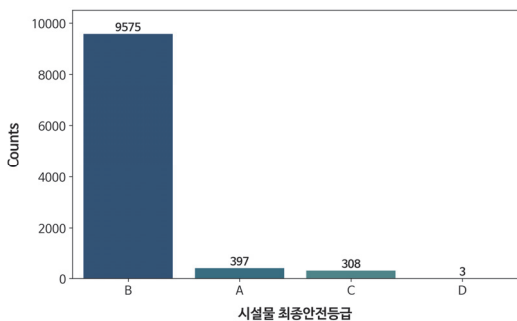


Fig. 2 Statistics on the facility safety rating

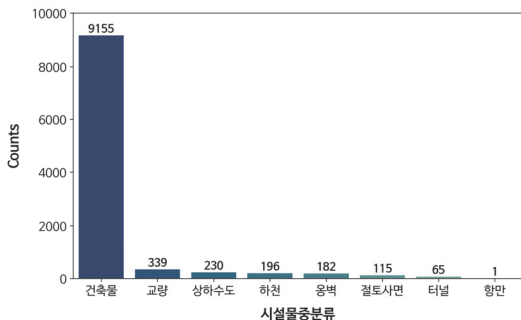


Fig. 3 Statistics on the facility classification

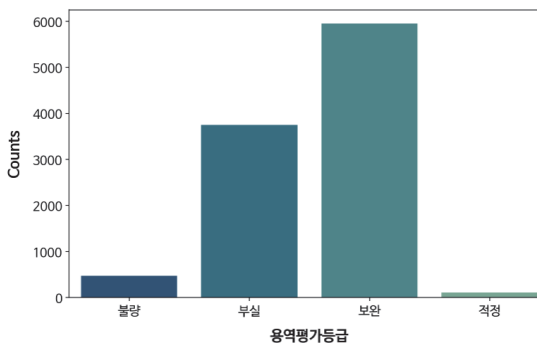


Fig. 4 Statistics on the evaluation rating

한 기술통계를 제시한다. 전처리 과정을 거친 10,283건 중 다음 변수들에 대한 분포 현황을 시각화하였다. “시설물 최종안전등급”, “시설물 중분류”, “평가등급”은 범주형 데이터이며, 각 변수의 분포는 Fig. 2부터 Fig. 4에 시각화해 나타내었다. “용역계약금액(log)”, “적정대가 준수율” 및 “용역 수행 일수” 변수는 수치형 데이터로 이에 대한 기술통계는 Table 3에 정리해 나타내었다. “시설물 최종안전등급”의 경우 대다수의 시설물에 대해 B 등급이 부여되었음을 확인할 수 있었다. 안전등급 B등급은 “보조 부재에 경미한 결함이 발생하였으나 기능 발휘에는 지장이 없으며, 내구성 증진을 위하여 일부의 보수가 필요한 상태[19]”로 용역 평가 등급과 별개로 시설물에 대한 안전은 확보된 상태임을 의미한다. “시설물중분류”의 경우 대다수의 점검·진단 대상 시설물이 건축물 형태로 나타났다. “용역 평가등급”은 타겟 변수로, 대부분이 부실 또는 보완 판정을 받았으며 전반적으로 클래스 간 불균형이 매우 높은 것으로 나타났다. 이는 데이터 수집 시 부실이 우려되는 점검·진단 보고서를 우선으로 선정했기에 ‘적정’ 사례는 낮은 분포를 나타낸다.

Table 3에 의하면 “용역계약금액_천원(log)”은 이미 로그 변환이 완료된 수치로 평균은 8.083 그리고 표준편차는 1로 나타났다. “적정대가 준수율”의 평균값은 0.420으로 대부분의 점검·진단이 적정기준 이하 가격선에서 수행되고 있음을 파악

Table 3. Descriptive statistics for numerical data

변수명	평균	표준편차	최소값	최대값	왜도
용역계약 금액_천원 (log)	8.083	1	5.328	11.016	0.818
적정대가 준수율	0.420	0.687	0.010	7	3.655
용역 수행 일수	27.365	24.913	0	424	4.007

할 수 있었다. “용역 수행 일수” 또한 점점 별 편차가 상당히 큰 것으로 나타났다. 특히 최소값과 최대값 사이의 간극이 매우 크며, 3가지 변수에 대한 왜도(skewness) 값이 모두 0보다 높음을 확인할 수 있었다. 왜도는 데이터 분포의 비대칭성을 나타내는 지표로, 분석 결과에 의하면 대다수는 평균에 가깝게 분포하지만, 극단적으로 큰 수치들이 그래프의 오른쪽에 치우쳐져 있음을 시사한다.

4.2 예측 모델 성능 비교

본 연구에서는 최적의 성능을 보이는 모델을 선정하기 위하여 다중 분류를 위한 5개의 머신러닝 알고리즘을 비교하였다. 학습 데이터와 검증 데이터는 7:3의 비율로 나누어 학습하였다. Table 4는 3가지 서로 다른 Scaling 기법과 SMOTE를 통해 처리한 입력 데이터의 성능지표를 Accuracy, Precision, Recall 및 F1-Score 순으로 정리해 나타낸다. 그 결과 SVM을 제외한 나머지 4가지 알

고리즘에 대한 편차는 크지 않음을 확인하였다. SVM은 분석 데이터가 선형적으로 분리 가능하다는 가정하에 작동하기 때문에 비선형 관계를 제대로 포착하지 못하며, 이진 분류를 기반으로 하기에 다중 클래스로 확장에 있어 복잡성이 증가하는 한계를 지닌다.

AdaBoost는 미세하지만 F1-score가 타 알고리즘 대비 낮은 특징을 보여, 일차 필터링에서는 SVM과 AdaBoost를 제외한 나머지 알고리즘을 선택하였다. 3개의 서로 다른 Scaler를 활용해 실험을 진행한 결과, 대체로 Min Max Scaler 적용 시 F1-score가 평균적으로 더 높게 나타났다. 또한, F1-score의 경우 Min Max Scaler와 RandomForest를 결합한 경우 가장 높게 나타났다.

4.3 최종 모델 선정

앞서 수행된 결과를 바탕으로 전체적인 성능이 우수하며 다중 클래스 분류 문제에 폭넓게 활용되

Table 4. Performance comparison among selected machine learning algorithms

Scaling	알고리즘	Accuracy	Precision	Recall	F1-score
Standard Scaler	XGBoost	0.9465	0.9431	0.9339	0.9385
	Random Forest	0.9562	0.9496	0.9367	0.9430
	LightGBM	0.9429	0.9551	0.9303	0.9422
	SVM	0.7319	0.5695	0.8186	0.6123
	AdaBoost	0.9524	0.9226	0.9453	0.9336
Min Max Scaler	XGBoost	0.9452	0.9508	0.9322	0.9413
	Random Forest	0.9572	0.9637	0.9375	0.9500
	LightGBM	0.9468	0.9618	0.9326	0.9466
	SVM	0.5556	0.4673	0.7024	0.4563
	AdaBoost	0.9501	0.9094	0.9340	0.9212
Robust Scaler	XGBoost	0.9491	0.9631	0.9337	0.9478
	Random Forest	0.9556	0.9505	0.9351	0.9426
	LightGBM	0.9417	0.9542	0.9283	0.9407
	SVM	0.7303	0.5564	0.7927	0.5943
	AdaBoost	0.9485	0.9238	0.9301	0.9269

Table 5. Grid search criteria for hyperparameter tuning

모델	파라미터	값 범위	선택된 값
Random Forest	n_estimators	{50, 100, 150, 200, 250}	150
	min_samples_split	{2, 3, 4, 5, 6}	5
	min_samples_leaf	{1, 2, 3, 4, 5}	1
	max_features	{'auto', 'sqrt', 'log2'}	log2
	max_depth	{23, 25, 27, 29, 31}	23
	criterion	{'gini', 'entropy'}	entropy
	class_weight	{none, 'balanced'}	None
	bootstrap	{true, false}	False
XGBoost	n_estimators	{50, 100, 150, 200, 250}	200
	reg_lambda	{0.8, 1.0, 1.2, 1.4, 1.6}	1.0
	reg_alpha	{0.01, 0.02, 0.03, 0.04, 0.05}	0.04
	max_depth	{11, 13, 15, 17, 19}	19
	learning_rate	{0.01, 0.02, 0.03, 0.04, 0.05}	0.04
	colsample_bytree	{0.4, 0.5, 0.6, 0.7, 0.8}	0.7
Light GBM	n_estimators	{50, 100, 150, 200, 250}	200
	max_depth	{3, 5, 7, 10, 13, 15}	15
	learning_rate	{0.01, 0.05, 0.1, 0.2}	0.2
	num_leaves	{31, 50, 70, 100}	100
	min_child_samples	{20, 30, 40, 50}	30
	subsample	{0.8, 0.9, 1.0}	0.8
	colsample_bytree	{0.8, 0.9, 1.0}	1.0
	reg_alpha	{0, 0.1, 0.5, 1.0}	0
reg_lambda	{0, 0.1, 0.5, 1.0}	1.0	

Table 6. Comparison of model performance after hyperparameter tuning

알고리즘	Accuracy	Precision	Recall	F1-score
XGBoost	0.9592	0.9653	0.9382	0.9512
Random Forest	0.9601	0.9704	0.9418	0.9555
LightGBM	0.9592	0.9652	0.9384	0.9512

는 Random Forest, XGBoost, LightGBM 세 모델에 대하여 하이퍼파라미터 튜닝을 실시하였다 [20]. 하이퍼파라미터 튜닝은 입력 가능한 하이퍼파라미터 값의 범위를 Grid에 사전 정의하고 모든 조합을 시도해 최적의 파라미터를 탐색하는 Grid Search 방식, 최적의 하이퍼파라미터를 찾기 위하여 탐색할 구간과 횟수를 정한 후 무작위로 값을 선택하는 Random Search 방식이 있다. 본 연구는 Grid Search 방식으로 찾아낼 수 없는 전체 영역에 대한 탐색이 가능하고, 검색 효율이 우수한 Random Search를 채택하였다. 스케일링 방식은 Min Max Scaler를 활용하였다. Table 5는 하이퍼파라미터 튜닝을 위해 탐색된 파라미터의 범위와 최종적으로 선택된 파라미터 값을 정리해 나타낸다. Table 6는 도출된 최적의 하이퍼파라미터 조합을 활용해 생성한 모델의 성능지표를 나타낸다.

하이퍼파라미터 튜닝을 수행한 결과 선택된 모든 모델의 성능지표가 소폭으로 상승하였다. 그중 Random Forest가 모든 성능지표에 가장 우수한 결과를 보였다. 이는 트리의 개수(n_estimators)와 최대 깊이(max_depth), 최대 특성 수(max_features) 등의 파라미터 조정이 모델 성능 향상에 큰 영향을 미친 것으로 나타난다. 트리의 개수는 랜덤 포레스트를 구성하는 결정 트리의 총 개수를 의미하며, 트리의 개수가 많아질수록 모델의 예측성능이 향상되는 경향을 보이지만 너무 많은 수의 트리는 과적합으로 이어질 수 있다. 마찬가지로 트리의 깊이가 깊어질수록 모델이 데이터의 복잡한 특성을 파악하기에 수월하지만, 과적합 위험은 증가할 수 있다. 최대 특성 수는 모델의 일반화 능력에 영향을 미치는 변수다. 본 연구에서 선택된 "log2"의 경우 상대적으로 낮은 값에 해당하며, 이는 개별 트리의 과적합 위험을 감소시키고, 트리 간의 상관성을 줄여 다양성을 증가시켜 모델의 성능향상에 기여한다.

4.4 결정요인 분석

앞의 과정을 통하여 하이퍼파라미터 최적화를 수행하였고 가장 성능이 우수한 Random Forest를 최종모델로 선정하였다. 이후 용역 등급평가에 영향을 미치는 결정요인을 식별하기 위하여 Permutation Importance 분석을 진행하였다. 해당 결과는 Fig. 5에 정리하였다.

변수의 중요도는 ‘용역계약금액_천원_(log)’, ‘용역수행일수’, ‘용역적정대가 준수율’, ‘기사 및 기술사 보유’, ‘고급·특급 기술인보유’, ‘석사 및 박사 보유’ 순으로 도출되었다. 이는 낮은 계약 금액을 앞세워 용역을 수주하는 것보다, 과업 수행에 적절한 계약금액을 확보하는 것이 점검·진단 용역의 품질과 높은 관련성이 있음을 시사한다. 아울러 점검·진단을 위해 소요되는 ‘용역 수행일수’ 역시 부실 용역 예방에 큰 영향을 미칠 수 있

음을 확인할 수 있었다. 특히, ‘용역일수’는 ‘용역 계약금액_천원_(log)’과도 상관성이 높기 때문에 적정 수행기간 확보가 중요함을 의미한다. 또한, 용역 수행업체의 기술력과 직결되는 ‘기사·기술사 보유’, ‘고급·특급 기술인 보유’, ‘석사 및 박사 보유’와 같은 변수가 모델의 예측성능에 있어 높은 중요도를 지닌 것으로 나타났다. 즉, 숙련된 인력의 보유는 용역 수행결과의 품질을 높이는 데 기여할 수 있지만, 관련 요인들이 낮거나 부적절하게 관리될 경우, 용역이 "부실"이나 "불량"으로 평가될 가능성이 높음을 시사한다.

5. 결론

시설물 관련 안전사고는 대규모 인명 피해와 재산 손실을 초래함에 따라 국민의 삶에 지대한 영향을 미친다. 더불어 노후 시설물의 급증은 안전 등급 유지에 부정적 영향을 끼치고 있어, 시설물의 현재 안전상태에 대한 객관적이고 투명한 점검이 시급한 실정이다. 국토안전관리원이 수행 중인 안전 점검·진단 용역 적정성 평가는 국민 생활 터전의 안전 사각지대 해소에 기여하고 있지만, 절차가 복잡하고 많은 행정력과 시간이 소요된다. 본 연구를 통하여 구축한 머신러닝 기반 용역 등급 평가 예측모델은 용역보고서 제출 즉시 분석·예측에 활용할 수 있어 신속한 의사결정에 기여할 수 있다. 본 연구에서는 점검·진단 용역의 부실 판정에 기인하는 요인을 발굴하기 위하여 기존 연구에서 다뤘던 시설물 관련 요인 외 용역 수행업체, 용역 관련 요인들을 포함하여 분석하였다. 비교한 알고리즘 중 Random Forest 기반 모델이 가장 우수한 성능을 나타내었고 하이퍼파라미터 조정으로 성능지표를 향상하였다. 이후 Permutation Importance 분석을 통해 용역평가 등급에 영향을

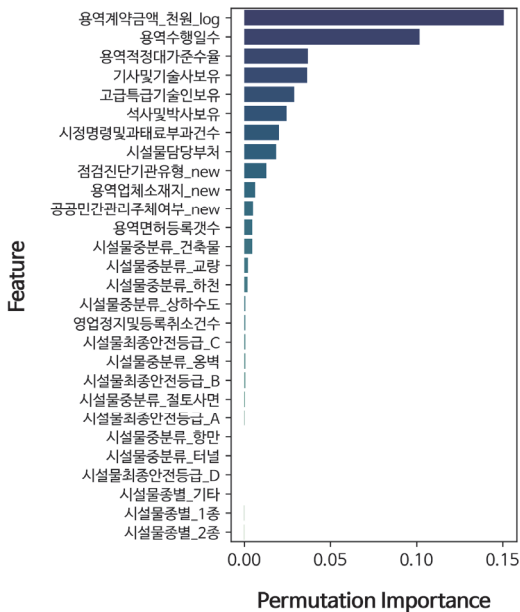


Fig. 5 Understanding the importance of input variable

미치는 결정요인을 도출하였다. ‘용역계약금액’, ‘용역수행일수’, ‘적정대가 준수율’이 가장 중요한 요인들로 도출되었고, 이는 용역에서 요구하는 과업을 원활하게 수행하기 위하여 적정한 용역금액과 충분한 과업일수 기획이 용역 품질과 깊은 연관성이 있음을 확인하였다. 또한, 기존 연구에서 다루지 않았던 용역업체 전체 인력 중 ‘기사·기술사’, ‘고급·특급 기술자’, ‘석·박사 학위 보유자’의 숙련자 비율도 중요성이 높은 것으로 확인되었다.

본 연구는 다음의 한계점을 지닌다. 본 연구에서는 정형 데이터로 분석 대상을 한정하였고 비정형 데이터인 ‘점검·진단 보고서 내용’, ‘보수·보강 필요 사항’ 같은 변수들은 포함하지 않았다. 자연어 처리 기법을 적용해 언급된 변수들로부터 유의미한 정보를 추출한다면, 이번 연구에서 규명하지 못했던 새로운 결정요인 또는 용역평가 등급별 보고서 작성 패턴 등을 파악하는 데 유용할 것으로 판단된다. 후속 연구에서는 텍스트마이닝 분석을 추가로 접목해 명시한 한계점을 보완해 국민의 안전을 위협하는 요인들을 더 세밀하게 규명하고자 한다.

사 사

본 논문은 산업통상자원부의 ‘융합기술사업화 확산형 전문인력 양성사업’의 지원을 받아 수행된 논문임.

참고문헌

- [1] 통계청. 2023, “근로자 10만명당 치명적 산업 재해수(OECD회원국)”, <https://kosis.kr/statHtml/statHtml.do?orgId=101&tblId=DT_2KAA308_OECD&conn_path=I2>, viewed 6 October (2023).
- [2] 고용노동부, 2022년 산업재해 현황 부가통계 (재해조사 대상 사망사고 발생 현황)[보도자료] <https://www.moel.go.kr/news/enews/report/enewsView.do?news_seq=14546>, viewed 19 January (2023).
- [3] 경기도 감사총괄담당관, 중대재해 예방 공중이용시설 등 안전관리 실태-도민감사관 참여 특정감사 결과 보고[사전정보공표], <https://www.gg.go.kr/opendata/openDataContent.do?gg_content_idx=9331&decodeIdx=7241>, p.3-4, (2023).
- [4] 하명호, 박종섭, 부실 정밀점검 및 정밀안전진단의 문제점과 평가제도의 개선방향, 한국구조물진단유지관리공학회 논문집(KSMI), 15(5), p.160-168, (2011).
- [5] 시설물의 안전 및 유지관리에 관한 지침 (2024. 4. 17., 국토교통부고시 제2024-202호) 제63조(평가의 대상)
- [6] 시설물의 안전 및 유지관리에 관한 특별법 (2021. 9. 17., 법률 제17946호) 제18조 제3호
- [7] 한범진, 김태희, 손승현, 건설 안전관리를 위한 Safety Climate 평가요인별 중요도 분석 연구, 한국건축시공학회지 논문집, 23(5), p.607-618, (2023).
- [8] 정인수, 소규모 시설물의 안전관리 강화방안, 한국안전학회지, 29(6), p111-118, (2014).
- [9] 이정민, 김율성, 토픽모델링을 활용한 항만안전 위험요인 도출에 관한 연구, 한국항만경제학회지, 39(2), p.59-76, (2023).
- [10] 김진국, 양충현, 박수빈, 텍스트 마이닝을 이용한 터널 교통안전 연구동향 분석, 디지털콘텐츠학회논문지, 23(10), p.2075-2083, (2022).
- [11] 이상규, 비정형 텍스트 기반의 토픽 모델링을 이용한 건설 안전사고 동향 분석. 한국산학기술학회 논문지, 19(10), p.176-182, (2018).
- [12] 장태연, 윤시후, 지석호, 임석빈. XGBoost를 활용한 시설물의 부재 상태 예측, 한국건설관리학회논문집, 24(1), p.31-39, (2023).
- [13] 홍지수, 전세진, 의사결정나무 및 랜덤포레스트 분류 모델을 이용한 교량 안전등급 예측, 대한토목학회논문집, 43(3), p.397-411, (2023).
- [14] 김정집, 임재운, 김성환, 정희경, 인공지능 기반 구조물 안전관리 예측 시스템. 한국지식정

- 보기술학회 논문지, 18(2), p.247-257, (2023).
- [15] Rousseeuw, P. J., Hubert, M., Robust statistics for outlier detection, WIREs: Data mining and knowledge discovery, 1(1), p.73-79. (2011).
- [16] Chawla, N. V., Bowyer, K. W., Hall, L. O., Kegelmeyer, W. P., SMOTE: synthetic minority over-sampling technique, Journal of Artificial Intelligence Research, 16, p.321-357, (2002).
- [17] 김성진, 안현철. 기업신용등급 예측을 위한 랜덤 포레스트의 응용, 산업혁신연구, 32(1), p.187-211, (2016).
- [18] Korobov, M., Lopuhi, K., "Welcome to ELI5's documentation!", <<https://eli5.readthedocs.io/en/latest/index.html>>, viewed 10 June (2024).
- [19] 시설물의 안전 및 유지관리에 관한 지침 (2024. 4. 17., 국토교통부고시 제2024-202호) 별표 12(시설물의 안전등급, 제29조 관련)
- [20] Bergstra, J., Bengio, Y., Random search for hyper-parameter optimization, Journal of Machine Learning Research, 13(2), p.281-305, (2012).

(접수: 2024.07.01. 수정: 2024.07.14. 게재확정: 2024.07.22.)