

서울교육 데이터 분석·활용 아이디어 공모전

– ‘디지털 전환을 통한 교육청 학교평가 자동화’ 기획안 –

목 차

1. 배경 및 목적	1
2. 분석내용	2
(1) 분석의 핵심 내용	2
(2) 분석 자료 및 데이터 수집	2
(3) 데이터 전처리	2
(4) 변수 및 지표 설명	2
(5) 분석 방법	3
3. 분석결과	4
(1) 상관분석	4
(2) anova 분석	4
(3) 모델설계	4
(4) 최종 학교평가예측모델 성능	5
4. 활용 방안 및 기대효과	6
(1) 활용방안 및 기대효과	6
(2) 분석의 한계점 및 개선안	6
5. 참고문헌	7

표 목차

[표 1] 변수 설명표	2
[표 2] 지표 설명표	3
[표 3] 그룹간 평균차이검정결과	4
[표 4] cross-validation를 통한 모델 성능 비교결과	5
[표 5] 최종 예측모델 성능	5

그림 목차

[그림 1] 2024 학교평가 가이드북 평가문항 예시	1
[그림 2] 변수간 상관관계 분석	4
[그림 3] 실제값 예측값 비교	5

1. 배경 및 목적

교육에 있어서 학생 평가만큼이나 중요하고 필요한 것이 바로 학교평가이다. 『2024 학교평가 평가지표』에서 그 필요성을 확인해 볼 수 있는데, 먼저 ‘교육의 질 향상’이다. 기본적으로 교육에 있어서 평가는 더 나은 교수·학습 환경을 제공하고 지속적으로 문제가 되는 상황들을 개선하는 것에 있다고 할 수 있다. 두 번째는 ‘책임성과 투명성 강화’이다. 학교평가는 학교 운영에 가담하는 대상들의 책임감을 더욱 고양할 수 있을 것이며 교육목표를 잘 달성하고 있는가에 대한 투명한 척도로 작용할 수 있다. 세 번째는 ‘학생 안전 및 복지 증진’이다. 안전하고 쾌적한 교육 환경을 조성하는 것은 교육수요자에게 있어서 교육의 권리를 행사할 기준으로 작용하기도 한다. 네 번째는 ‘학교 발전 방향 제시’이다. 학교평가의 척도를 통해 상대적인 부분 혹은 절대적인 부분으로 나타난 부분에 대한 반성과 발전 방향을 모색하고 더욱 개선된 학교 환경을 만들어낼 수 있다. 다섯 번째는 ‘학부모와 지역사회의 신뢰 확보’이다. 평가를 통해 산출된 척도가 함께 교육하는 지역사회, 학부모에 대한 신뢰를 얻을 수 있는 명분이 생기며 교육수요자의 알권리 역시 충족시킬 수 있게 된다. 이렇듯 학교평가는 교육수요자의 알권리, 더욱 개선된 교수학습 방법을 위해 필요한 상황이지만, 실제로는 다음과 같은 문제점들로 인해 평가의 어려움이 있다.

2013년부터 학교평가 중 외부 평가를 폐지하고 학교 자체평가로 전환하고자 하는 노력이 시작되어, 서울특별시교육청은 2014년부터 학교평가 중 외부 평가를 폐지했다. 하지만 그 결과는 학교 자체평가를 잘 진행하지 않고 있으며, 그 원인을 파악해 본 결과 “교원 업무 부담 증가”가 가장 큰 원인으로 작용하는 것을 확인할 수 있었다.

평가지표	평가문항
학교 비전 공유 및 실현	<ul style="list-style-type: none"> • 학교 교육공동체의 합의를 바탕으로 학교 비전을 도출하였다. • 학교의 교육목표를 구체적으로 정하고 이를 달성하기 위해 노력하였다. • 학교평가 결과를 반영하여 학교교육계획을 수립하였다.
학교자율운영체제 내실화	<ul style="list-style-type: none"> • 학교 교육공동체의 의견수렴 체계가 활성화되어 있다. • 학교평가 결과 공유 및 개선 방안 마련을 위해 토론회 등을 체계적으로 추진하였다. • 학교교육목표 달성을 위해 교육활동 중심의 학교평가 평가지표를 개발하여 활용하였다. • 학교 교육공동체의 요구를 바탕으로 학교자율 사업운영제를 내실 있게 운영하였다.

<그림 1 : 2024 학교평가 가이드북 평가문항 예시>

또한 현 학교평가는 정성적인 평가에 치우쳐 있어 정량적으로 데이터 분석을 하고 평가점수를 산출하는 것이 쉬운 일은 아닌 상황이다. 따라서 본 공모전에서는 서울시 학교별 데이터를 이용하여 “교육수요자의 알권리 충족”을 위한 정량적 지표의 개발과 객관적인 데이터를 통한 자동화된 평가점수 시스템을 구축하여 “교원 업무 부담 경감”을 목표로 한다.

2. 분석 내용

(1) 분석의 핵심 내용

‘서울 열린 데이터 광장’에서 제공하는 공공데이터의 분석을 통해 교육수요자의 요구 충족, 교육정책의 이행된 정도, 교원 업무 부담 경감을 위한 학교평가 AI 모델 생성, 추상적인 개념의 수치화 및 점수화를 목적으로 한다.

(2) 분석 자료 및 데이터 수집

본 공모전에서 사용된 데이터는 모두 ‘서울특별시교육청’에서 제공하는 ‘서울시 학교별’ 데이터 총 22개의 csv 형식 파일을 활용하였으며, 활용한 데이터 목록은 별첨하였다.

(3) 데이터 전처리

수집한 데이터 목록에서 순회학급 학급당 학생 수, 특수학급 학급당 학생 수의 결측치는 0으로 채웠으며, 학교 시설 개방 여부에 데이터의 결측치들은 ‘시설없음’으로 채웠다. 또한, 데이터 셋을 합치는 과정에서 발생한 변수들의 결측치들은 히스토그램과 상자 그림을 활용하여 분포 양상에 따라 결측치를 채워 최종 데이터 셋을 구축하였다.

(4) 변수 및 지표 설명

본 공모전에서는 서울특별시교육청교육연구정보원에서 공개한 2024 학교평가 평가지표 및 평가문항 자료를 참고하여, 아래의 변수들을 활용해 크게 네 가지 범주에서 평가할 수 있는 지표를 만들었다.

변수	변수 설명	변수	변수 설명
X_1	일반교과교실 개방 여부	X_{16}	응급처치교육 시간
X_2	특별교실 개방 여부	X_{17}	장애인 주 출입구 접근로 설치 여부
X_4	수업교원 1인당 학생 수	X_{19}	장애인 주 출입구 높이 차이 제거 여부
...
X_5	학생 100명당 동아리 수	X_{20}	장애인 출입구(문) 설치 유무
X_{12}	폭력예방및신변보호교육 시간	X_{27}	경보 및 피난설비 설치 여부
X_{13}	약물중독예방교육 시간	X_{28}	체육장 개방 여부
X_{14}	지능정보서비스과의존관련교육 시간	X_{29}	체육관 개방 여부
X_{15}	직업안전교육 시간	X_{30}	강당 개방 여부

< 표 1 : 변수 설명표 >

지표	지표명	산출방법
S_1	교육시설 평가점수	$S_1 = a_1 + a_2$ $a_1 = x_1 + x_2 + x_3 + x_{28} + x_{29} + x_{30}$ $a_2 = \alpha \times \log(x_9)$ <p> $- x_9$: 학생 100명당 교육시설 수로 로그값을 취한 후 더함 $- \alpha$: x_9을 얼마나 반영할 것인지(본 분석에서는 0.5를 적용) </p>
S_2	교육과정운영 평가점수	$S_2 = (b_1 + b_2) - b_3$ $b_1 = x_6 + x_7 + x_8$ $b_2 = \beta \times \log(x_5)$ $b_3 = \delta \times \log(x_4)$ <p> $- x_5$: 학생 100명당 동아리 수로 로그값을 취한 후 더함 $- \beta$: x_5을 얼마나 반영할 것인지(본 분석에서는 0.7을 적용) $- x_4$: 수업교원 1인당 학생 수, 업무 과중성에 대한 패널티 척도를 위해 로그값을 취한 후 더함 $- \delta$: 업무 과중 패널티를 얼마나 반영할 것인지 (본 분석에서는 0.05를 적용) </p>
S_3	안전교육시간 이수 평가점수	$S_3 = \sum_{i=10}^{16} x_i$
S_4	베리어프리 점수	$S_4 = \sum_{i=17}^{27} x_i$
Y	최종학교평가 점수	$y = s_1 + s_2 + \log(s_3) + \log(s_4)$

< 표 2 : 지표 설명표 >

(5) 분석 방법

본 공모전에서는 python을 활용하여 데이터 수집, 전처리, 통계분석, AI 모델링을 수행하였다. 활용한 통계분석은 변수간 상관분석과 학교평가점수 상위 그룹과 하위 그룹간 평균점수 차이 검정을 수행하였다. 또한, 자동으로 학교평가를 할 수 있는 AI 예측 모델을 설계하기 위해 cross-validation 기법을 활용하여 학습 데이터에서 가장 뛰어난 예측 성능을 보인 모델을 선정하였고, 선정된 모델을 다양한 파라미터 조합으로 튜닝하여 최종 예측 모델을 설계하였다.

3. 분석결과

(1) 상관분석



< 그림 2 : 변수간 상관관계 분석 >

변수간 상관관계 분석결과 생활안전교육시간(X_{10}) 교통안전교육시간(X_{11})이 가장 높은 연관성을 보이고 있다. 또한, 베리어프리 변수, 교육시설평가에 관한 변수 간에도 비교적 높은 연관성을 보인다. 이는 각 카테고리에서 사용된 변수 간에는 어느 정도 연관성이 있음을 나타낸다.

(2) anova 분석

일원배치분산분석인 anova 분석은 2개 이상의 그룹 간 평균의 차이 검정을 위한 통계분석 방법이다. 산출된 최종 학교 평가 점수를 기준으로 총 4개의 그룹으로 나누어 네 가지 범주(S_i)의 점수에 대한 평균차이검정을 수행하였다. anova 분석을 통해 통계적 유의성을 확인한 결과 유의수준 0.05에서 모두 통계적으로 유의미한 차이를 보인다.

지표	지표명	검정통계량	유의확률($p-value$)
S_1	교육시설 평가점수	3774.519	0
S_2	교육과정운영 평가점수	299.4378	5.99E-180
S_3	안전교육시간이수 평가점수	245.9176	1.14E-149
S_4	베리어프리 점수	56.8107	3.73E-36
Y	최종학교평가점수	7469.491	0

< 표 3 : 그룹간 평균차이검정결과 >

(3) 모델 설계

학교평가를 자동으로 수행하는 AI예측 모델을 설계하기 위해 학습 데이터와 평가 데이터를 7:3 비율로 나눈 후, 의사결정나무, 랜덤포레스트, lightgbm, xgboost, svm모형을 cross-validation 기법을 활용해 학습 데이터에 대한 예측 성능을 비교하였다. 비교 결과 lightgbm 모델의 예측 성능이 가장 뛰어났고, 해당 모델을 그리

드 서치 기법을 활용해 최적의 파라미터 조합인 'learning_rate': 0.1, 'max_depth': 5, 'n_estimator': 100을 찾아 최종 예측 모델을 설계하였다.

model	cv score
Decision Tree	0.908
Random Forest	0.959
XgBoost	0.969
Lightgbm	0.973
SVM	0.913

< 표 4 : cross-validation를 통한 모델 성능 비교결과 >

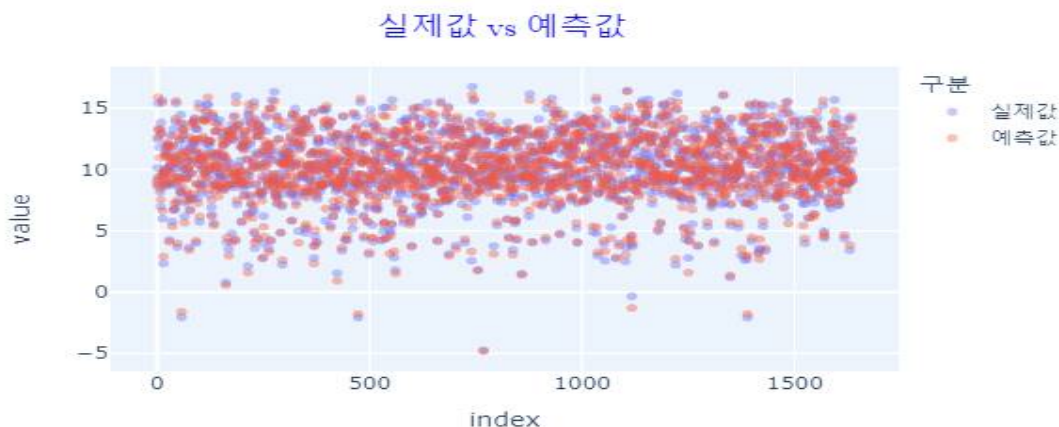
(4) 최종 학교평가예측모델 성능

평가 데이터를 적용하여 설계한 예측모델의 성능을 평가한 결과는 아래 표와 같다.

평가척도	산출방식	값
평균절대오차	$MAE = \frac{1}{N} \sum_i y_{pred} - y_{true} $	0.331
평균제곱오차	$MSE = \frac{1}{N} \sum_i (y_{pred} - y_{true})^2$	0.184

< 표 5 : 최종 예측모델 성능 >

각 평가척도에서 0.331, 0.184로 설계한 예측 모델이 평가 데이터에서 뛰어난 성능을 보이고 있음을 확인하였다. 또한, 아래 그림에서 볼 수 있듯이 각각의 개체에 대한 예측값이 실제값과 거의 일치함을 보인다.



< 그림 2 : 실제값 예측값 비교 >

4. 활용 방안 및 기대효과

(1) 활용방안 및 기대효과

첫째, 교육 현장에서 발생하는 데이터를 기반으로 효율적인 의사결정이 가능하다. 학교에서 일어나고 있는 모든 의사결정은 사람 간의 토의와 대화를 통해 주로 이루어지는데, 이는 자칫하면 어느 한 개인의 의사가 과하게 반영된 의사결정일 수 있다. 그러나 이러한 평가 자동화 모델이 예측 결과를 도출해 준다면 객관적이고 효율적인 의사결정이 가능할 것이다.

둘째, 단위 학교의 고유한 데이터가 쌓이게 되면 평가 자동화 모델을 통해 개별적인 학교 평가가 가능해진다. 또한, 매년 실시한 데이터 분석 내용이 축적되면 학교 평가 만족도 예측 패턴을 더 효과적으로 이해할 수 있으며, 학교만의 장기적인 교육계획 수립에 도움이 될 것이며, 또한 자동화된 평가로 인해 교원의 업무 역시 줄어들 것으로 기대된다.

셋째, 정성적인 평가에 치우쳐져 있던 기존 평가지표에서 정량적인 평가가 가능하게 된다. 여러 가지 요인들을 종합하여 분석한 결과 실제 계산한 값과 모델이 예측한 값 사이에 차이를 크게 보이지 않는 뛰어난 성능을 보이고 있음을 확인할 수 있었다. 따라서 해당 평가 자동화 모델과 수집되는 수치적인 데이터들을 통해 정량적인 평가를 진행할 수 있게 됐고, 기존의 주관적이고 정성적인 평가에서 정량적이고 객관적인 평가를 할 수 있게 될 것이며 이는 교원의 업무부담 경감이라는 효과를 가져올 것으로 기대된다.

(2) 분석의 한계점 및 개선안

분석에 있어서 존재했던 그 개선안은 총 두 가지로 요약된다. ‘학교활동에서 수집하는 데이터의 종류의 다양화’, ‘기존에 진행했던 정성적인 평가점수에 대한 데이터’의 제공이다.

우선 학교활동의 다양성을 고려한다는 것은 교육비 측면에서 ‘목적사업비’에 대한 고려를 하겠다는 의미로 이는 교육여건의 도출로 이어지게 된다. 목적사업비와 지역의 교육여건의 상관이 존재하기 때문에, 교육의 여건까지 반영된 더욱 공신력이 있는 평가지표가 되기 위해서는 학교활동에서 수집하는 데이터의 다양화가 필요하다. 다음으로 본 공모전에서 도출된 학교평가점수가 실제 그 기능을 하는가에 관한 더욱 높은 타당도를 확보하기 위해서는 기존에 진행했던 평가점수에 대한 데이터의 제공이다. 이를 통해 학교평가점수를 개발하는 과정에서 더욱 상관이 높은 변수를 찾아내고, 지표를 개발하여 확인된 결과가 실제 그 역할을 하는지, 정성적인 평가방식을 대체할 수 있는지에 관한 더욱 확실한 근거를 제시할 수 있게 될 것이다.

5. 참고문헌

- [1] 김주아(2006). 교육성과를 포함한 책무성 중심 학교평가 모형 개발 및 적용. 한국교육평가학회
- [2] 김영식(2021). 시도교육청별 학생수 및 교육비특별회계 세출결산액 변화 분석. 한국교육재정경제학회
- [3] 송기창&윤홍주(2014). 일반고등학교의 교육재정운영 및 교육비 결정요인에 대한 설립별 분석. 한국교육재정경제학회
- [4] 강상진&황정원(2010). 학교 책무성 체제에서 학업성취도에 기초한 학교 성과 지표는 중등학교 학교평가 지표로서 적합한가?: 학교효과 이론에 의한 타당도 검증. 한국교육학회
- [5] 장현진(2019). 일반계 고등학생의 진로성숙도에 대한 개인, 가정, 학교 및 진로활동 경험의 영향. 서울교육대학교 석사학위논문
- [6] 류민석(2016). 『학교자체평가 효과성 제고 방안』.서울교육
- [7] 『학교회계 예산편성 기본지침, 학교행정업무메뉴얼』.(2024).교육부
- [8] 『2024 학교평가 평가지표』.(2024).서울특별시교육청교육연구정보원

서울교육 데이터 분석·활용 아이디어 공모전 데이터 분석 기획보고서(일반부)

작품명	※참가신청서 작품명과 동일하게 작성	
주제 분야	<input type="checkbox"/> 교육과정-수업평가 <input type="checkbox"/> 고교다양화 <input type="checkbox"/> 체육·예술교육 <input type="checkbox"/> 유아교육 <input type="checkbox"/> 기회의평등 <input type="checkbox"/> 민주시민·인권교육 <input checked="" type="checkbox"/> 학교업무경감·교원정책 <input type="checkbox"/> 서울형혁신학교 및 미래교육지구(구.혁신교육지구) <input type="checkbox"/> 안전 <input checked="" type="checkbox"/> 서울교육공동체 <input type="checkbox"/> 교육공간 변화	
활용 데이터 목록		
제공 기관(사이트) 명	데이터명	출처(URL)
서울 열린데이터광장	서울시 학교별 교육운영 특색사업 계획 현황	https://data.seoul.go.kr/dataList/OA-22333/S/1/datasetView.do
서울 열린데이터광장	서울시 학교별 보건관리 현황	https://data.seoul.go.kr/dataList/OA-22337/S/1/datasetView.do
서울 열린데이터광장	서울시 학교별 직원 현황	https://data.seoul.go.kr/dataList/OA-22344/S/1/datasetView.do
서울 열린데이터광장	서울시 학교별·학급별 학생수 현황	https://data.seoul.go.kr/dataList/OA-22316/S/1/datasetView.do
서울 열린데이터광장	서울시 학교별 장애인 편의시설 현황	https://data.seoul.go.kr/dataList/OA-22334/S/1/datasetView.do
서울 열린데이터광장	서울시 학교별 안전교육 계획 및 실시현황	https://data.seoul.go.kr/dataList/OA-22338/S/1/datasetView.do
서울 열린데이터광장	서울시 학교별 대상별 학교폭력 예방교육 실적현황	https://data.seoul.go.kr/dataList/OA-22321/S/1/datasetView.do
서울 열린데이터광장	서울시 학교별 자유학기제 운영에 관한 사항 현황	https://data.seoul.go.kr/dataList/OA-22313/S/1/datasetView.do
서울 열린데이터광장	서울시 학교별 방과후학교 운영 계획 및 운영 지원현황	https://data.seoul.go.kr/dataList/OA-22342/S/1/datasetView.do
서울 열린데이터광장	서울시 학교별 교사 현황	https://data.seoul.go.kr/dataList/OA-22324/S/1/datasetView.do
서울 열린데이터광장	서울시 학교별 입학생 현황	https://data.seoul.go.kr/dataList/OA-22322/S/1/datasetView.do

서울 열린데이터광장	서울시 학교별 동아리활동 현황	https://data.seoul.go.kr/dataList/OA-22340/S/1/datasetView.do
서울 열린데이터광장	서울시 학교별 학생의 체력 증진에 관한 사항	https://data.seoul.go.kr/dataList/OA-22345/S/1/datasetView.do
서울 열린데이터광장	서울시 학교별 학교시설 개방에 관한 사항	https://data.seoul.go.kr/dataList/OA-22326/S/1/datasetView.do
서울 열린데이터광장	서울시 학교별 표시과목별 교원 현황	https://data.seoul.go.kr/dataList/OA-22320/S/1/datasetView.do
서울 열린데이터광장	서울시 학교별 성별 학생수 현황	https://data.seoul.go.kr/dataList/OA-22315/S/1/datasetView.do
서울 열린데이터광장	서울시 학교별 방과후학교 운영 계획 및 운영 지원현황	https://data.seoul.go.kr/dataList/OA-22342/S/1/datasetView.do
서울 열린데이터광장	서울시 학교별 교사 현황	https://data.seoul.go.kr/dataList/OA-22324/S/1/datasetView.do
서울 열린데이터광장	서울시 학교별 입학생 현황	https://data.seoul.go.kr/dataList/OA-22322/S/1/datasetView.do
서울 열린데이터광장	서울시 학교별 동아리 활동 현황	https://data.seoul.go.kr/dataList/OA-22340/S/1/datasetView.do
서울 열린데이터광장	서울시 학교별 학생 학부모 상담계획 및 실시 현황	https://data.seoul.go.kr/dataList/OA-22343/S/1/datasetView.do
서울 열린데이터광장	서울시 학교별 교복 구매 유형 및 단가 현황	https://data.seoul.go.kr/dataList/OA-22332/S/1/datasetView.do