

<http://dx.doi.org/10.17703/JCCT.2024.10.5.667>

JCCT 2024-9-78

## 공무원 직무 전문교육 만족도 예측을 위한 딥러닝 기반 회귀 모델 설계

### A Deep Learning-based Regression Model for Predicting Government Officer Education Satisfaction

오수민\*, 윤성연\*, 박민서\*\*

Sumin Oh\*, Sungyeon Yoon\*, Minseo Park\*\*

**요약** 공직자로서의 바람직한 가치관 확립과 공직의 전문성 향상을 위해 공무원의 직무 전문교육이 강조되고 있다. 만족도 높은 맞춤형 직무교육을 제공하기 위해 만족도에 영향을 미치는 요인을 분석한 연구들이 제안되고 있으나, 교육 내용을 활용하여 만족도를 예측한 연구는 부족한 실정이다. 따라서 본 연구는 교육 내용을 함께 고려해 공무원 직무 전문교육 만족도를 예측하는 딥러닝(Deep Learning) 모델을 제안한다. 제안 방법은 공무원 전문 교육과정 정보 데이터를 활용한다. 우선 문자형으로 수집된 변수인 교육 대상, 교육 구분, 교육 형태를 원-핫 인코딩(One-hot Encoding)으로 카테고리화(Categorized)한다. 교육을 통해 학습할 수 있는 내용이 문자형으로 저장된 교육 내용을 TF-IDF(Term Frequency-Inverse Document Frequency)으로 수치화한다. 이를 딥러닝 기반의 회귀 모델로 학습하고, 10-겹 교차 검증(10-Fold Cross Validation)으로 모델의 성능을 검증한다. 본 연구의 제안 모델은 테스트 데이터에서 99.87%의 높은 예측 정확도를 보인다. 향후 본 연구를 고려한 맞춤형 교육 추천은 교육 대상에 최적화된 교육을 제공 및 개선하는 데에 도움이 될 것으로 기대한다.

**주요어** : 공무원 직무 전문교육, 교육 만족도, 딥러닝, 회귀 모델

**Abstract** Professional job training for government officers emphasizes establishing desirable values as public officials and improving professionalism in public service. To provide customized education, some studies are analyzed factors affecting education satisfaction. However, there is a lack of research predicting education satisfaction with educational contents. Therefore, we propose a deep learning-based regression model that predicts government officer education satisfaction with educational contents. We use education information data for government officer. We use one-hot encoding to categorize variables collected in text format, such as education targets, education classifications, and education types. We quantify the education contents stored in text format as TF-IDF. We train our deep learning-based regression model and validate model performance with 10-Fold Cross Validation. Our proposed model showed 99.87% accuracy on test sets. We expect that customized education recommendations based on our model will help provide and improve optimized education content.

**Key words** : Government officer education, Education satisfaction, Deep learning, Regression

\*준회원, 서울여자대학교 데이터사이언스학과 대학원생

\*\*정회원, 서울여자대학교 데이터사이언스학과 조교수  
(교신저자)

접수일: 2024년 5월 21일, 수정완료일: 2024년 6월 20일

게재확정일: 2024년 9월 1일

Received: May 21, 2024 / Revised: June 20, 2024

Accepted: September 1, 2024

\*\*Corresponding Author: mpark@swu.ac.kr

Dept. of Data Science, Seoul Women's Univ, Korea

## I. 서론

성공적인 사회적 경제 발전과 산업재해 예방을 위해 직무별 전문인력 양성이 필요하다[1]. 국가통계포털(Korean Statistical Information Service, KOSIS)이 공개한 ‘데이터 전문인력 양성을 위한 정책적 지원사항’에 따르면, 정부는 관련 학과 개설, 지속적인 교육 제공 등 직무별 전문인력 양성을 위한 다양한 정책 및 교육을 시행하고 있다[2]. 또한, 인적자원개발위원회(Industrial Skills Council, ISC)는 인력 수요가 높은 직무를 중심으로 훈련, 자격 등의 기준을 마련해 인력양성을 지원하고 있다[3].

정부의 인력양성과 다양한 교육의 양적 확대 운영에도 불구하고 현장 적합도가 높은 맞춤형 인력양성은 여전히 부족하다[4]. 이에, 한국자산공사(Korea Asset Management Corporation, KAMCO) 인재개발원은 신규 인력의 직무수행을 위한 기본과정과 실제 업무 적용이 가능한 집중과정, 전문성 강화를 위한 심화 과정으로 교육체계를 세분화하여 업무 경험과 이해도에 따른 맞춤 교육을 제공하고 있다[5].

이와 함께 직무 전문교육의 질과 만족도를 개선하기 위해, 교육 만족도에 영향을 미치는 요인을 분석한 연구가 활발하게 진행되고 있다[6-9]. 그러나, 기존 연구는 교육 내용을 고려하지 않고 기타 교육 관련 변수를 기반으로 직무 전문교육 만족도의 결정 요인을 탐색하는 한계가 존재한다. 교육 내용은 교육 만족도에 직접적인 영향을 미치기 때문에, 직무 전문교육 만족도 분석 시 함께 고려해야 한다.

최근 공직자로서의 바람직한 가치관 확립과 공직의 전문성 향상을 위해 공무원의 직무별 전문교육이 특히 강조되고 있다[10]. 2022년 11월 시행된 공무원 인재개발법 제 9조에 따라, 공무원 직무 전문교육 내용은 지속적으로 평가 및 개선되고 있다[11].

이에, 본 연구에서는 교육 내용을 함께 고려한 공무원 직무 전문교육 만족도 예측을 위한 딥러닝(Deep Learning) 모델을 제안하고자 한다. 공무원 직무 전문교육과 관련된 변수와 교육 내용 데이터를 수집한 후, 딥러닝 기반의 회귀(Regression) 모델을 개발한다. K-겹 교차 검증(K-Fold Cross Validation)을 통해 모델의 성능을 검증한다.

본 논문은 다음과 같이 구성된다. 제 2장에서는 교육

만족도에 영향을 미치는 요인에 대해 분석한 선행 연구를 살펴본다. 제 3장에서는 본 연구에서 제안한 딥러닝 기반의 회귀 모델을 개발한다. 제 4장에서는 연구 결과를 설명한다. 제 5장에서는 본 논문의 결론을 언급한다.

## II. 관련 연구

본 장에서는 직무 전문교육 만족도에 영향을 미치는 요인을 분석한 선행 연구를 살펴본다.

Ho 외[6]는 홍콩 대학 학부생 423명을 대상으로 수집한 성별, 소속 교수진, 교육 접근성 등 27개의 변수를 활용하여 코로나19 기간 동안 시행한 긴급 원격 학습에 대한 학생 만족도를 예측하였다. 학습 방법이 가장 중요한 예측 변수였으며, 강사의 노력 수준, 평가 방법에 대한 합의 등이 원격 학습을 통한 전공교육 만족도에 긍정적인 영향을 미치는 것을 확인하였다.

Sahin[7]은 터키 아나톨리아 대학교의 917명을 대상으로 수집한 학생 만족도, 강사의 도움, 학생 상호작용 및 협업, 개인적 관련성, 실제 학습, 적극적인 학습 및 학생 자율성 등 6개 변수를 활용하여 대학 개설 교과의 교육 만족도에 미치는 영향을 분석하였다. 개인적 관련성, 강사의 도움, 적극적인 학습, 실제 학습 순으로 개설 교과의 교육 만족도에 긍정적인 영향을 미치는 것을 증명하였다.

한예진[8]은 한의과대 재학생 372명을 대상으로 수집한 응답자의 기본 특성, 교육과정의 만족도, 교수자의 만족도 등 6개의 영역에서 총 53개의 변수를 활용하여 한의과대 학생을 위한 전공 교과 교육의 전체 만족도에 영향을 미치는 변수를 분석하였다. 교육과정, 교육평가, 교육환경이 교육 전체 만족도에 긍정적인 영향을 미치며, 교수자의 만족도는 유의미한 영향을 미치지 않는 것을 증명하였다.

김혜정[9]은 대학생 100명을 대상으로 줌(Zoom)을 활용한 온라인(Online) 기반 전문 강의의 효율성을 파악하기 위해 학습자의 만족도를 조사 및 분석하였다. 줌을 활용한 온라인 강의는 학습에 대한 흥미와 동기 부여에 도움을 주며, 자기주도학습과 적극적 상호작용이 가능하다는 특성은 강의 만족도에 긍정적인 영향을 미치는 것을 확인하였다. 그러나, 기계적 오류나 결함, 음질 불량, 효율적인 수업을 위해 부가적으로 사용해야 하는 맞춤 기능 등은 강의 만족도에 부정적인 영향을

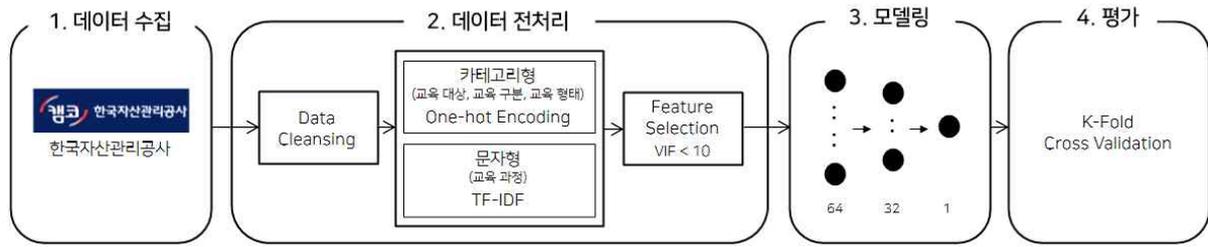


그림 1. 공무원 직무 전문교육 만족도 예측을 위한 딥러닝 모델링  
 Figure 1. Flow diagram of deep learning-based government officer education satisfaction prediction model

미치는 것을 확인하였다.

이처럼 선행 연구는 직무 전문교육 만족도에 영향을 미치는 요인을 분석하는 데에 초점을 맞춰 진행되고 있다. 그러나 직무별 맞춤형 전문교육의 고도화를 위해 교육 만족도를 분석하기 위해서는 교육을 통해 학습할 수 있는 내용이 포함될 필요가 있다. 따라서 본 논문에서는 교육 내용을 함께 고려한 직무 전문교육 만족도를 예측하는 딥러닝(Deep Learning) 모델을 제안한다.

### III. 연구방법

본 연구는 데이터 수집, 데이터 전처리, 딥러닝 기반 교육 만족도 예측 모델 개발 세 단계로 구성된다. 본 연구의 프로세스는 그림 1과 같다.

#### 1. 데이터 수집

한국자산관리공사(Korea Asset Management Corporation, KAMCO)에서 제공하는 교육과정 정보 데이터 중, 교육 만족도에 영향을 끼치는 요인을 수집한다. 만

표 1. 독립변수의 분산팽창지수  
 Table 1. Variance Inflation Factors(VIF) of Independent variables

독립변수	Variance Inflation Factor
교육 형태_오프라인	7.2
교육 형태_온라인	3.1
교육 구분_직무공통교육	2.5
교육 형태_온오프라인 병행	2.3
교육 형태_실시간 온라인	1.7
수료 인원	1.6
강사 수	1.5
교육 구분_정보화 교육	1.5
교육 대상	1.5
교육 내용_TF-IDF	1.5
교육 형태_온라인	1.3
교육 차수	1.2

족도 예측을 위해 교육 구분, 교육 차수, 교육 형태, 교육 내용, 교육 대상, 교육 시간, 수료 인원, 강사 수 등 8개 변수로 구성된 264개 데이터를 활용한다. 해당 데이터는 수치형 변수 4개와 문자형 변수 4개로 구성되어 있다.

#### 2. 데이터 전처리

수집한 데이터 중, 변수 일부가 수집되지 않은 53개의 데이터는 결측치로 판단하여 제거한다. 문자형으로 저장된 데이터 중, 카테고리(Category)로 분류 가능한 교육 대상, 교육 구분, 교육 형태 등 3개의 변수를 수치형으로 변환한다. 원-핫 인코딩(One-hot Encoding)을 활용해 교육 대상 2개, 교육 구분은 4개, 교육 형태는 5개의 카테고리로 구분한다.

교육 내용은 교육을 통해 학습할 수 있는 내용이 문자형으로 저장된 변수이다. 교육 내용의 정보를 기반으로 수치형으로 변환하기 위해 명사 추출, 토큰화(Tokenization), TF-IDF(Term Frequency - Inverse Document Frequency)를 수행한다. 교육 내용에 포함된 모든 내용을 형태소의 단위로 분리한다. 명사(noun)를 추출하여 토큰화(Tokenization)를 수행한다. 토큰화한 교육 내용 내의 모든 명사를 TF-IDF를 활용해 수치화한다. TF-IDF는 특정 문서에서 특정 단어가 얼마나 중요한지를 나타내는 값이다[12]. TF-IDF 값이 클수록 교육 내용의 주제가 타 교육에 비해 특화되었다는 것을 의미한다. TF-IDF를 활용하여 비슷한 교육 내용을 포함하고 있으나 교육 과정명은 다른 유사한 교육을 그룹화할 수 있다는 장점이 있다. 본 연구에서는, 각 교육 내용에 포함되는 모든 토큰의 TF-IDF 값을 추출한 뒤, 이를 모두 더해 교육 내용을 수치화하였다.

전처리 한 모든 변수를 분산팽창지수(Variance Inflation Factor, VIF)를 활용하여 다중공선성 기준 유

의성을 평가한다. VIF의 값이 10 이상인 경우, 해당 변수가 다른 변수 간 높은 상관관계를 가진다고 판단한다. 이를 기반으로 유의한 독립변수를 도출한다. VIF 결과는 표 1과 같다.

### 3. 모델링

데이터 전처리 후 도출된 16개의 변수로 구성된 데이터를 교육 만족도 예측을 위한 딥러닝 모델에 적용한다.

실험을 위하여 전체 데이터에서 훈련 데이터(Training Sets)와 테스트 데이터(Test Sets)의 비율을 각각 80%, 20%로 나누어 구성한다.

실험에 사용된 모델은 입력층과 2개의 은닉층, 출력층으로 구성된다. 입력층에 입력된 데이터는 64개, 32개의 노드로 구성된 은닉층을 거쳐 출력층을 통해 예측값을 출력한다. 은닉층의 활성화 함수로는 ReLU (Rectified Linear Unit)를 사용하고, 출력층의 활성화 함수로 Linear를 사용한다. 제안 네트워크의 하이퍼파라미터(Hyper-parameter)인 배치(Batch) 크기는 16으로 설정한다. 손실함수는 MSE(Mean Squared Error)를 사용하고, 최적화 함수는 Adam Optimizer를 사용한다. 총 100번의 반복 학습을 수행하고, 예측 모델의 정확도를 검증하기 위해 10-겹 교차 검증(10-Fold Cross Validation)을 수행한다.

## IV. 실험 및 결과 분석

제안 모델의 설명력 평가 및 검증을 위한 성능 평가 지표로 결정 계수(R-Squared)와 평균 제곱 오차(Mean Squared Error, MSE), 정확도(Accuracy)를 측정하였다. R-Squared는 모델의 설명력을 나타내며, MSE는 모델의 예측력을 나타낸다. 표 2는 본 연구에서 제안한 딥러닝 기반의 공무원 직무 전문교육 만족도 예측 모델의 정량적 평가 결과이다.

R-Squared 값이 1에 가까울수록 설명력이 좋다고 판단하며, MSE 값이 작을수록 실제값과 예측값 간의 오차가 적은 모델이므로 예측 정확도가 높다고 해석할 수 있다. 본 연구의 실험 결과는 표 2와 같이, 테스트 데이터에서 R-Squared 0.9975, MSE 0.0079, Accuracy 99.87%의 우수한 성능을 보이는 것을 알 수 있으며, 10-겹 교차 검증(10-Fold Cross Validation) 결과 역시 R-Squared 0.9992, MSE 0.0030, Accuracy 99.92%로

표 2. 공무원 직무 전문교육 만족도 예측 모델의 정량적 평가  
Table 2. Quantitative evaluation of government officer education satisfaction prediction model

데이터 세트	Evaluation		
	R-squared	MSE	Accuracy
Training Sets	0.9993	0.0026	99.96%
10-Fold Cross Validation	0.9992	0.0030	99.92%
Test Sets	0.9975	0.0079	99.87%

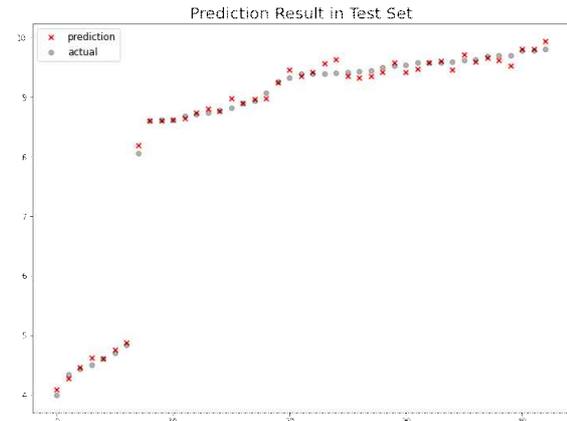


그림 2. 공무원 직무 전문교육 만족도 예측 결과 그래프: 빨간색 X는 예측값을, 회색 O는 실제값을 의미

Figure 2. Government officer education satisfaction prediction result graph: Red X means predicted value and gray O means actual value

안정적인 성능을 보이는 것을 알 수 있다.

그림 2는 제안한 모델의 성능을 보여주는 그래프로, 테스트 데이터를 기반으로 예측값과 실제값의 차이를 나타낸다. 그래프의 빨간색 X는 예측값, 회색 O는 실제값을 의미한다. x-축은 랜덤하게 구성된 테스트 데이터를 실제값 기준으로 오름차순 정렬하여 새롭게 부여한 Index이고, y-축은 각 Index에 해당하는 교육 만족도를 의미한다. 이 그래프를 통해 모델은 실제값과 유사한 추세로 예측하는 것을 알 수 있다.

## V. 결론

본 논문에서는 딥러닝(Deep Learning)을 활용하여 공무원의 직무 전문교육 만족도를 예측하는 딥러닝(Deep Learning) 모델을 제안하였다. 공무원 직무 전문교육 관련 변수와 교육 내용 데이터를 활용하였다. 변수 일부가 수집되지 않아 모델의 예측력에 부정적인 영향을 끼칠 수 있는 결측치를 제거하였다. 교육 대상, 교육 구분, 교육 형태 변수를 원-핫 인코딩(One-hot Encoding)

을 활용하여 카테고리화(Categorized)하였다. 교육 내용 변수를 TF-IDF(Term Frequency - Inverse Document Frequency)를 활용하여 수치화하였다. 전처리를 거친 변수 중, 다중공선성 기준 유의한 변수를 탐색하기 위해 분산팽창지수(Variance Inflation Factor, VIF)를 기준으로 유의성을 평가하였다. 최종 선정된 변수를 활용하여 딥러닝 모델에 적용하고, 10-겹 교차 검증(10-Fold Cross Validation)을 통해 검증하였다. 실험 결과, 제안한 공무원 직무 전문교육 만족도 예측 모델은 테스트 데이터(Test Sets) 기준 99.87%의 정확도(Accuracy)로, 우수한 성능을 가짐을 알 수 있었다.

본 연구에서 제안한 공무원 직무 전문교육 만족도 예측을 고려한 맞춤형 교육 제공은 이전 회차 교육의 만족도를 반영하여 교육생에게 최적화된 맞춤형 교육을 제공하는 데 도움이 될 것이라 기대한다.

그러나, 연구에 사용한 데이터의 양이 적고 만족도의 분포가 극명하게 나뉘어 있어 데이터 품질 향상이 필요하다는 한계가 존재한다. 향후 연구에서는 데이터 증강(Data Augmentation)을 통해 데이터의 양을 늘리고, 교육 내용에 담긴 키워드를 카테고리로 분류하여 교육 내용을 더욱 명확하게 제시하고자 한다.

## References

- [1] Seoul Social Economy Center. Available online: <https://sehub.net> (accessed on 20 June 2024)
- [2] Korean Statistical Information Service (KOSIS). Available online: <https://kosis.kr> (accessed on 20 June 2024)
- [3] Industrial Skills Council (ISC). Available online: <https://www.isckorea.or.kr> (accessed on 20 June 2024)
- [4] J. Lee & H. Jang, "Current Status and Policy Issues of Collaborations between Universities and Family Companies in Korea", *Journal of the Economic Geographical Society of Korea*, Vol. 23, No. 1, pp. 71-81, 2020. DOI: 10.23841/egsk.20.23.1.71
- [5] Korea Asset Management Corporation (KAMCO) Human Resources Development Institute. Available online: <https://hrd.kamco.or.kr> (accessed on 20 June 2024)
- [6] I.M.K. Ho, K.Y. Cheong, and A. Weldon, "Predicting student satisfaction of emergency remote learning in higher education during COVID-19 using machine learning techniques", *PLOS ONE*, Vol. 16, No. 4, April 2021. DOI: 10.1371/journal.pone.0249423
- [7] I. Sahin, "Predicting student satisfaction in distance education and learning environments", *Turkish Online Journal of Distance Education (TOJDE)*, Vol. 8, No. 2, pp. 113-119, April 2007.
- [8] Y. Han, "A Study on the Factors affecting Korean Medical Students' Satisfaction with Education", *The Journal of the Convergence on Culture Technology (JCCT)*, Vol. 10, No. 3, pp. 253-258, 2024. DOI: 10.17703/JCCT.2024.10.3.253
- [9] H. Kim, "The Efficacy of Zoom Technology as an Educational Tool for English Reading Comprehension Achievement in EFL Classroom", *International Journal of Advanced Culture Technology (IJACT)*, Vol. 8, No.3, pp. 198-205, 2020. DOI: 10.17703/IJACT.2020.8.3.198
- [10] T. Cho, "Exploratory Analysis on Strengthening Public Professionalism: Focusing on Institution for Improving Public Professionalism", *Institute of Public Policy and Administration*, Vol. 34, No. 1, pp 101-129, March 2020. DOI: 10.17327/ippa.2020.34.1.005
- [11] Korea Law Information Center. Available Online: <https://www.law.go.kr> (accessed on 20 June 2024)
- [12] A. Akiko, "An information-theoretic perspective of tf-idf measures", *Information Processing & Management*, Vol. 39, No. 1, pp. 45-65, 2003. DOI: 10.1016/S0306-4573(02)00021-3