

설명 가능한 인공지능을 활용한 퍼스트 마일에서의 트럭 운송 임금 산정

이샘미[†], 강경태^{*}

[†]한양대학교 인공지능융합학과, ^{*}한양대학교 인공지능학과

[†]saemmi@hanyang.ac.kr, ^{*}ktkang@hanyang.ac.kr

Truck Transport Wage Calculation in First Mile Using Explainable AI

Saemmi Lee[†], Kyungtae Kang^{*}

[†]Dept. of Applied Artificial Intelligence, Hanyang University

^{*}Dept. of Artificial Intelligence, Hanyang University

요약

퍼스트 마일에서의 트럭 운전자 임금은 물류 기업의 관리자가 운송 거리, 운전사의 경력, 당시의 임금 시세를 바탕으로 결정한다. 이러한 방식은 관리자의 주관적인 판단이 개입되어 임금 차별을 초래할 수 있으며, 관리자와 운전자 간의 상호 확인을 위한 추가적인 절차가 필요하다. 본 논문에서는 이와 같은 문제를 해결하기 위해, 설명 가능한 인공지능을 활용하여 운전자 임금을 예측하는 방법을 제안한다. 제안된 모델의 코사인 유사도는 0.9017, 결정 계수는 0.7122로 계산되었으며, 예측된 임금에 대한 근거를 제공하여 임금 산정 결과에 대한 투명성과 이해도를 높여 줄 것으로 기대된다.

I. 서론

퍼스트 마일(First Mile)이란 물류 네트워크의 시작점으로서, 생산 시설이나 제조 공장에서 상품이 출발하여 다른 물류 시설로 운송되는 과정을 말하며, 이 과정에서 트럭 운전사는 핵심적인 역할을 수행한다[1]. 이러한 트럭 운전사의 임금은 운송 거리, 운전사의 경력, 당시의 임금 시세를 바탕으로 물류 기업에서 결정되고 상호 확인을 거친 뒤 지급된다.

현재의 임금 산정 시스템은 물류 기업의 관리자가 조건을 기반으로 임금을 결정한다. 그러나 이러한 방식은 불공정하거나 주관적인 판단으로 임금 차별이 발생할 수 있으며, 일관성이 부족할 수 있다. 또한 임금 결정에 대한 동의와 이해를 얻기 위해 관리자와 운전자 간의 상호 확인 및 조율을 위한 보완적인 절차가 필요하다.

본 연구에서는 위와 같은 문제를 해결하기 위해, 설명 가능한 인공지능을 활용하여 운송 거리와 운전자 정보, 임금 시세를 기반으로 트럭 운전사의 임금을 예측하고 이에 대한 근거를 제공한다. 이는 임금 산정 자동화의 기반이 되어, 관리자의 주관적인 판단을 배제할 수 있으며, 해당 작업에 필요했던 인력과 시간을 감소시키면서 기업에는 경제적인 혜택을 제공한다. 또한 관리자와 운전사에게 임금 산정 결과에 대한 이해를 높여 상호 확인을 위한 절차를 최소화할 수 있다.

그림 1은 연구의 아키텍처를 나타낸 것으로, 데이터 추출부터 전처리, 모델 생성 및 평가 과정을 도식화했다. 해당 과정을 거친 트럭 운전자 임금 예측 모델은 예측된 임금과 더불어 임금 도출에 대한 근거를 제공할 수 있으며, 코사인 유사도는 0.9017, 결정 계수는 0.7122로 계산되었다.

II. 제안 기법

2.1 데이터 추출 및 전처리

학습을 위해, 미국 소재의 물류 회사 시스템에서 트럭 운전자 데이터와 2023년 1월부터 6월까지의 임금 데이터를 추출하였다. 트럭 운전자 데이

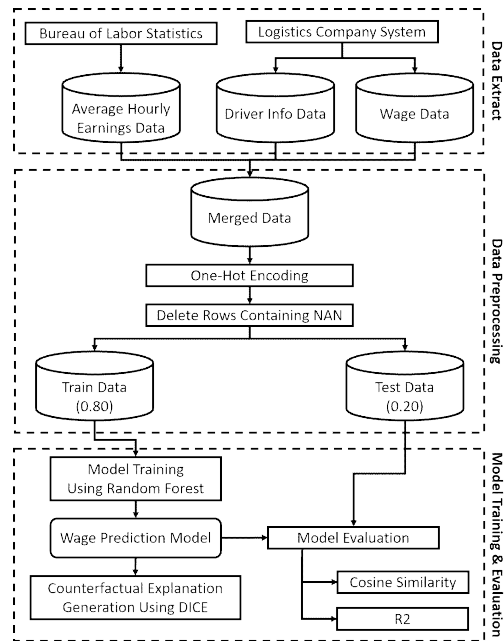


그림 1. 아키텍처

터에는 운전사의 고유 ID 및 경력, 유형 정보가 담겨 있었으며, 임금 데이터에는 각 운전사의 고유 ID, 지급된 임금, 운송 거리의 범주에 따라 분류된 코드 그리고 지급일 정보가 포함되어 있었다. 추가로 임금 시세를 모델에 반영하기 위해 미국 노동 통계국(Bureau of Labor Statistics)[2]에서 월별 시간당 평균 임금 데이터를 추출하였다.

데이터 결합을 위해 먼저 운전자와 임금 데이터에 공통으로 존재하는 운전자 고유 ID를 기준으로 두 데이터를 병합하였으며, 이후에는 임금 지급일을 기준으로 월별 시간당 평균 임금 데이터를 병합하였다. 이렇게 생성된 병합 데이터에서 학습에 사용할 feature와 target은 표 1과 같다.

* Corresponding Author

† Major in Bio Artificial Intelligence

표 3. 예측된 임금과 목표 임금에 대한 반사실적 설명

(a) Predicted Result

EARNINGS	EXPERIENCE	TYPE_EMP	TYPE_OWN	TYPE_CAR	CODE_5102	CODE_5103	...	CODE_5176	PRED
33.54	0.0	1	0	0	0	0	...	0	180.00

(b) Counterfactual Explanation for Target Wage

EARNINGS	EXPERIENCE	TYPE_EMP	TYPE_OWN	TYPE_CAR	CODE_5102	CODE_5103	...	CODE_5176	PRED
33.54	4.8	0	1	0	0	1	...	0	208.70

표 1. 학습에 사용할 feature와 target

	Column	Description
Feature	experience	운전자 경력
	type	운전자 유형
	code	운송 거리의 범주에 따라 분류된 코드
	earnings	시간당 평균 임금
Target	wage	운전자 임금

표 2. 코사인 유사도와 결정 계수

Cosine Similarity	0.9017
Coefficient of Determination	0.7122

속변수를 잘 설명한다. 생성된 모델의 결정 계수는 0.7122이다.

데이터 병합 후, 모델이 원활하게 패턴을 파악할 수 있도록 결측값이 포함된 행을 제거하였으며, 범주형 데이터인 type과 code에 대해서는 One-Hot Encoding을 진행하였다. 더불어 experience와 earnings는 최솟값과 최댓값이 동적으로 변하는 데이터이므로, 스케일링을 적용할 경우 데이터의 특성을 왜곡시킬 우려가 있어 원래의 데이터 그대로 학습시켰다.

2.2 Random Forest

Random Forest[3]는 복수 개의 Decision Tree를 생성하고, 각 Decision Tree가 개별적으로 학습을 진행한 뒤 결과를 취합하여 최종 예측값을 도출하는 방법이다. 해당 알고리즘은 비선형적 관계를 포착하고 모든 변수 간의 스케일 차이에 크게 영향을 받지 않는 특징이 있어, 별도로 스케일링을 적용하지 않은 전처리 과정을 고려하여 해당 알고리즘을 선택하였다.

본 연구에서는 Sklearn에서 제공되는 RandomForestRegressor 클래스와 전처리가 완료된 훈련 데이터를 활용하여 임금 예측 모델을 생성하였다.

2.3 DiCE

DiCE(Diverse Counterfactual Explanations)[4]는 설명 가능한 인공지능 분야에서 사용되는 알고리즘 중 하나이며, 해당 알고리즘은 모델의 예측을 설명하기 위해 반사실적 설명을 제공한다. 반사실적 설명이란 예측 결과를 설명하기 위해 해당 결과를 바꾸기 위한 대안적인 사례를 제공하는 것이다. 이는 운전자 임금에 영향을 미치는 요소를 이해하고, 모델이 어떤 조건에서 높은 혹은 낮은 임금을 예측하는지 파악할 수 있다.

본 연구에서는 예측된 값에 대한 설명을 생성하기 위해 사전에 훈련된 모델과 오픈 소스 라이브러리인 DiCE를 활용하였다.

III. 실험 결과

3.1 임금 예측 모델에 대한 평가

학습된 모델 평가를 위해 테스트 데이터를 사용하였다. 예측된 임금과 실제 임금으로 코사인 유사도(Cosine Similarity)와 결정 계수(Coefficient of Determination)[5]를 계산하였으며, 그 결과는 표 2와 같다.

코사인 유사도는 두 벡터 간의 코사인 각도를 이용해 유사성을 측정하는 방법이다. -1 이상 1 이하의 값을 가지며 1에 가까울수록 유사도가 높아진다. 생성된 모델의 코사인 유사도는 0.9017이다.

결정 계수는 회귀 모델에서 종속변수의 분산 중에서 독립변수로 설명되는 비율을 말한다. 0 이상 1 이하의 값을 가지며 1에 가까울수록 모델이 종

3.2 예측된 임금에 대한 설명 생성

예시로 180달러로 예측된 임금에 대해 200~210달러의 목표 임금을 설정하고 설명을 생성하였다. 표 3의 (a)는 독립 변수에 의해 예측된 임금이며, (b)는 설정된 목표 임금을 달성하기 위한 조건을 제시한 모습이다.

IV. 결론

본 연구는 운송 거리 및 운전자 정보, 임금 시세를 활용하여 임금 예측 모델을 개발하고, 설명 가능한 인공지능 기법을 도입하여 임금 산정 과정의 투명성을 높이고 관련 절차를 최소화하였다. 본 연구에서는 운송 거리의 범주에 따라 분류된 코드를 활용하여 모델을 학습시켰으나, 향후 연구에서는 운송의 시작과 종료 지점 간의 자세한 거리 정보를 활용한다면 더욱 정밀한 모델을 구현할 수 있을 것으로 기대된다.

ACKNOWLEDGMENT

이 논문은 2024년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원(No.RS-2022-00155885, 인공지능융합혁신인재양성(한양대학교 ERICA))과 2024년도 정부(산업통상자원부)의 재원으로 한국산업기술포럼의 지원을 받아 수행된 연구임(P0012744, 2024년 산업혁신인재성장지원사업)

참고 문헌

- [1] S. Lee and K. Kang, 'Automating Wage Calculation for First-Mile Truck Transportation Using Linear Regression', Korea Software Congress, 2023.
- [2] United States Bureau of Labor Statistics, 'Average Hourly Earnings', [Online]. Available: <https://www.bls.gov>. [Accessed: May, 3, 2024]
- [3] Ho, Tin Kam, 'Random decision forests', in Proceedings of 3rd international conference on document analysis and recognition, vol. 1, pp. 278-282, 1995.
- [4] Wachter S., Mittelstadt B. and Russell C., 'Counterfactual explanations without opening the black box: Automated decisions and the GDPR', Harv. JL & Tech., vol. 31, p. 841, 2017.
- [5] Lewis-Beck C. and Lewis-Beck M., 'Applied regression: An introduction', Sage publications, vol. 22, 2015.