

CatBoost 기반 트럭 운송 지연 예측에 대한 연구

차민재[†], 강경태^{*}

[†] 한양대학교 인공지능융합학과, ^{*}한양대학교 인공지능학과

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

Predicting Truck Transportation Delays Using CatBoost

Minjae Cha[†], Kyungtae Kang^{*}

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

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

요약

기술 혁신으로 인한 물류 산업 시장의 성장은 운송 지연 관리의 중요성을 증가시키고 있다. 본 논문에서는 트럭 운송 지연을 예측하기 위해 CatBoost를 기반으로 한 모델을 제안한다. 실험 결과, 제시한 모델은 98.83%의 정확도로 트럭 운송 지연 여부를 예측하였다. 추가적으로 이 연구에서는 다른 기계 학습 알고리즘과 비교하여 제시한 모델의 성능을 검증한다. 이 비교 분석을 통해 CatBoost 기반 모델이 다른 모델들보다 높은 성능을 보이는 것을 확인하였다.

I. 서론

물류 산업은 기술 혁신을 통해 지속해서 발전하고 있으며, 이는 효율적인 운송 관리의 중요성을 증대시키고 있다. 특히 트럭 기반 운송 업계에 종사하는 인력 또한 증가하는 추세이다 [1]. 이러한 배경하에서, 운송 지연은 공급망의 비용과 성능에 중대한 영향을 미치는 주요 요소이다. 따라서, 운송 지연 여부를 정확하고 신속하게 예측할 수 있는 능력은 물류 운영의 최적화와 비즈니스 프로세스 효율성 향상에 결정적인 역할을 수행한다.

이 연구는 물류 트럭 운송 지연 여부를 예측하여 운송 지연으로 인해 발생할 수 있는 문제 상황을 사전에 대응하는 것을 목표로 한다. 이를 위해 CatBoost [2] 알고리즘을 사용한 모델을 제안한다. 이 모델은 트럭 운송 데이터를 학습하여 운송 과정에서 발생할 수 있는 다방면의 원인을 분석하고 지연 예측 결과를 제시한다.

그림 1은 이 연구에서 제시하는 운송 지연 예측 모델의 아키텍처이다. 제안하는 모델의 성능을 향상시키기 위해 결측 값 제거, 표준화, 데이터 조정과 같은 최적화 과정을 수행한다. 이 과정 이후 CatBoost 알고리즘을 활용하여 모델을 학습시킨 뒤 검증을 거쳐 운송 지연 여부를 예측한다.

실험 결과, 제안한 기법을 통해 운송 지연 여부를 예측할 수 있었다. 더해서, 세 종류의 다른 기계 학습 알고리즘과 결과를 비교하여 제안한 모델의 성능을 검증하였다.

검증 결과, CatBoost 기반 모델이 다른 기계 학습 알고리즘보다 높은 정확도로 운송 지연 여부를 예측함을 알 수 있었다. 이러한 결과는 논문에서 제안한 모델이 실제 물류 운송에 있어서 발생할 수 있는 지연을 예측하여 예상치 못한 지연에 효과적으로 대응할 수 있고, 운영에서의 의사결정 과정을 효율적으로 지원할 수 있음을 시사한다.

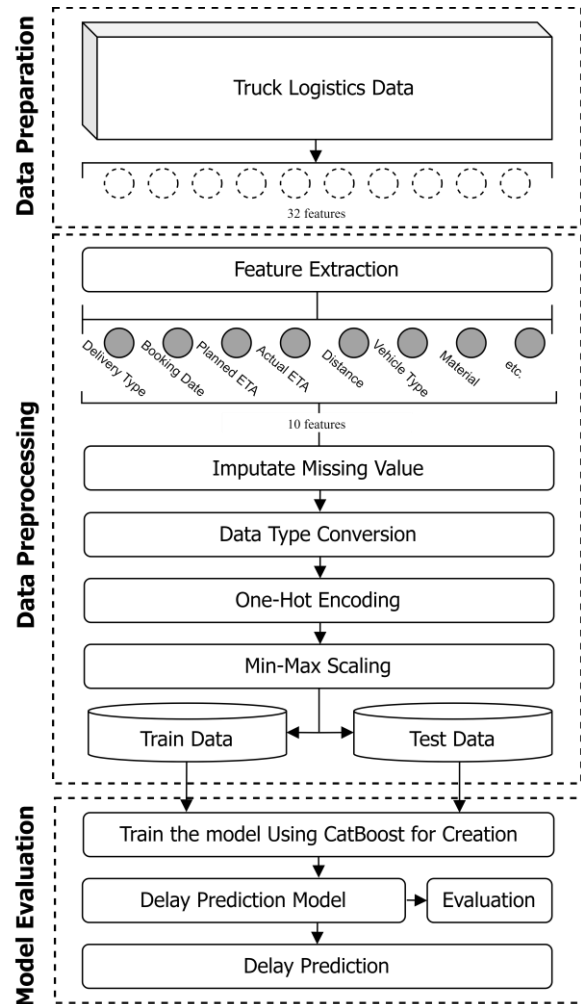


그림 1. 지연 예측 모델 아키텍처

* Corresponding Author

[†] Major in Bio Artificial Intelligence

표 1. 운송 데이터의 주요 특성

	Column	Description
Feature	market	미계약 공급업체
	regular	계약 공급업체
	transport_distance	운송 거리
	min_covered	최소 운행 거리
	gap_eta	도착 예정 시간차
	gap_bk_planned	도착 예정, 계약 시간차
	gap_bk_actual	실제 도착, 계약 시간차
	vehicle_type	트럭 유형
	material	운송 물품
Target	delay	운송 지연 여부

표 2. 모델 간 Confusion Matrix 의 정확도 비교

Model	FN	FP	TN	TP	Acc.
Linear Regression	76	191	304	803	80.56%
Decision Tree	178	47	448	701	83.62%
Random Forest	5	59	436	874	95.34%
CatBoost	9	7	488	870	98.83%

II. 본 론

본 연구는 Kaggle 에서 제공한 트럭 운송 기록 데이터 [3]를 사용하여 모델을 학습 및 검증한다. 이 데이터는 6,880 건의 운송 데이터 및 운송 형태, 출발지, 배송 도착지 등을 포함한 32 개의 항목으로 이루어져 있다. 모델의 성능을 향상시키기 위해 데이터 전처리를 진행하였다. 물류 운송을 수행하는 트럭의 종류와 운송의 대상이 되는 물품 항목의 정확한 수치 비교를 위해 One-Hot Encoding 을 수행하여 수치 데이터로 변환하였으며, 이 외의 날짜와 같은 문자열 데이터 또한 수치화를 진행하였다.

표 1 은 전처리 완료 데이터로 총 10 개의 특성으로 구성되어 있으며 예측하고자 하는 값은 Target 인 운송 지연 여부이고, Feature 는 Target 을 예측하는 데 사용되는 9 개의 특성이다. 데이터 유형 별로 할당될 수 있는 값의 범위가 상이하기에 특성 간의 일관된 규모를 확보하기 위해서 Min-Max Scaling [4]을 통해 값의 범위를 조정하였다. 연구에서 제안하는 모델의 일반화 능력을 평가하고 과적합을 방지하기 위해 데이터를 4:1 비율로 각각 훈련 데이터와 테스트 데이터로 분리하여 모델을 학습하고 평가하였다.

이 연구는 CatBoost 모델을 중심으로 모델 간 비교를 통해 트럭 운송 지연 예측에 가장 적합한 알고리즘을 식별하고자 하였다. 공정한 비교 기준을 제공하기 위해 동일한 데이터를 사용하여 각 모델을 훈련하였으며 모델의 테스트 결과 CatBoost 모델이 98.83%의 정확도로 다른 모델과 비교하여 우수한 성능을 보였다.

표 2 는 각 모델의 Confusion Matrix 분석 결과를 제시한 자료로 랜덤 포레스트 95.34%, 결정 트리 83.62%, 선형 회귀 80.56%의 정확도로 CatBoost 모델의 뒤를 이었다. 그림 2 의 AUC 값 또한 제안한 모델이 다른 모델에 비해 정확도가 높음을 나타낸다.

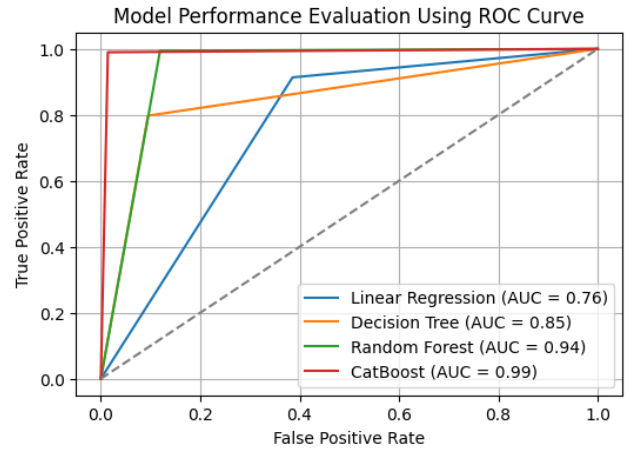


그림 2. ROC 곡선을 통한 예측 모델 평가

III. 결 론

본 논문은 CatBoost 기반의 트럭 운송 지연을 예측하는 모델을 제안한다. 실험 결과, 제안한 모델은 1,374 개의 테스트 지연 데이터 중 1,358 개의 지연 여부를 정확하게 예측하여 98.83%의 정확도를 나타냈다. 다른 모델과 비교 분석한 결과, CatBoost 모델은 다른 여러 알고리즘보다 높은 정확도를 보인 것을 확인하였다. 결과적으로, 제시한 모델은 운송 지연 여부를 예측하여 지연으로 인해 발생할 수 있는 문제를 사전에 대처할 수 있는 능력을 기르고 물류 운영과 공급망의 최적화와 같은 업무 효율성 향상에 큰 응용 가능성이 있을 것으로 기대된다.

ACKNOWLEDGMENT

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

참 고 문 헌

- [1] M. Cha and K. Kang, 'Predicting Truck Logistics Transport Delays Using Logistic Regression Analysis', Korea Software Congress, 2023.
- [2] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, 'CatBoost: unbiased boosting with categorical features', Advances in neural information processing systems, vol. 31, 2018.
- [3] Ram Thiagu, 'Delivery truck trips data - Logistics supply chain dataset', Aug 28, 2020. Distributed by Kaggle. <https://www.kaggle.com/datasets/ramakrishnanthiyagu/delivery-truck-trips-data>
- [4] V. N. G. Raju, K. P. Lakshmi, V. M. Jain, A. Kalidindi, and V. Padma, 'Study the Influence of Normalization/Transformation process on the Accuracy of Supervised Classification', in 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT), 2020, pp. 729- 735.