

Truck Weight Estimation using Operational Statistics at 3rd Party Logistics Environment

Yu-jin Lee* · Kyung Min Choi* · Song-eun Kim* · Kyungsu Park*[†] · Seung Hwan Jung**[†]

*Department of Business Administration, Pusan National University

**School of Business, Yonsei University

운영 데이터를 활용한 제3자 물류 환경에서의 배송 트럭 무게 예측

이유진* · 최경민* · 김송은* · 박경수*[†] · 정승환**[†]

*부산대학교 경영학과, **연세대학교 경영학과

Many manufacturers applying third party logistics (3PLs) have some challenges to increase their logistics efficiency. This study introduces an effort to estimate the weight of the delivery trucks provided by 3PL providers, which allows the manufacturer to package and load products in trailers in advance to reduce delivery time. The accuracy of the weigh estimation is more important due to the total weight regulation. This study uses not only the data from the company but also many general prediction variables such as weather, oil prices and population of destinations. In addition, operational statistics variables are developed to indicate the availabilities of the trucks in a specific weight category for each 3PL provider. The prediction model using XGBoost regressor and permutation feature importance method provides highly acceptable performance with MAPE of 2.785% and shows the effectiveness of the developed operational statistics variables.

Keywords : Third Party Logistics, Truck Weight Estimation, Operational Statistics, XGBoost Regressor, Permutation Feature Importance

1. 서 론

많은 기업은 경쟁우위, 원가우위 등의 달성을 위해 다양한 아웃소싱 전략을 펼치고 있다. 아웃소싱(Out Sourcing)이란 회사 내부에서 수행하던 업무를 외부 업체로 위탁하면서 운영 효율성을 추구하는 경영 전략이다. 특히 산업 현장에서는 생산한 제품을 소비자에게 운송하기 위해 자사의 운송수단을 사용하기도 하지만 많은 제조 기업은 공급 사슬 전체 혹은 일부를 전문적인 물류 업체에 위탁하는

제3자 물류를 활용하여 물류의 효율성 증대를 꾀하려 한다.

하지만 일부 경우에는 제3자 물류를 활용함으로써 물류에 대한 통제력을 상실하거나 효율성 증대를 위한 또다른 문제에 직면하기도 한다. 예를 들면, 미국 캘리포니아주의 도로교통법에서는 고속도로를 운행하는 트럭의 무게를 80,000lbs로 제한하고 있는데[13], 제조회사에서는 이러한 무게 제한에 맞추어 제품들을 패키징하려 한다. 만약 무게 제한을 초과하도록 제품을 패키징한다면 반복작업으로 인해 비용 증가 및 배송지연으로 이어지며, 무게 제한보다 너무 적은 제품을 패키징한다면 효율성 저하로 이어진다. 특히 제3자 물류에서 할당하는 배송 트럭의 가용 적재량 또는 무게 제한 내에서 적재 가능한 무게를 미리 제공하지

Received 28 November 2022; Finally Revised 10 December 2022;
Accepted 12 December 2022

[†] Corresponding Author : ks.park@pusan.ac.kr, seunghwan.jung@yonsei.ac.kr

못하는 경우에는 제품을 미리 패키징하여 물류의 효율성을 증대시키기 힘들다. 본 연구에서는 제3자 물류를 활용하는 미국의 한 맥주회사의 실제 사례 및 데이터를 바탕으로 할당되는 배송 트럭의 무게를 예측하는 연구를 진행하였다.

배송 트럭의 무게 제한 및 무게 예측을 통한 적절한 적재 용량은 트럭의 도로 교통안전에도 긍정적인 영향을 미친다. 과적 트럭은 도로교통상황에서 제동거리가 길어 긴급 상황 대처 능력이 떨어지고 동력 부족으로 오르막길에서 속도가 느려지는 상황을 야기한다고 알려져 있다[1]. 안전 문제 외에도 과적 트럭은 차량의 타이어, 브레이크 및 기타 구성요소의 마모와 손상을 가속화시키기 때문에 운송비의 상승도 있을 수 있다.

배송트럭의 무게 예측을 위해 본 연구는 XGBoost Regressor를 사용하였다. 맥주회사에서 관리되는 기존 데이터 외에도 유가, 날씨, 인구수 등의 예측 변수를 활용하려 하였으며, 특히 운영 데이터에 기반한 예측 변수(operational statistics parameter)를 생성하여 예측의 성능 향상을 꾀하였다. 즉, 각 물류회사에서 맥주회사에게 할당된 배송 트럭의 과거 기록과 최근 할당된 배송 트럭 기록을 바탕으로 특정 무게 그룹에 속하는 배송 트럭의 가용 확률을 생성하였다. 이러한 예측 변수들 중 유의미한 변수를 파악하기 위해 Permutation Feature Importance (PFI)와 통계적 검정을 활용하였다.

본 논문의 구성은 다음과 같다. 제2장에서는 관련된 기존 연구에 대해 소개하고, 제3장에서는 본 연구에서 다루는 문제 상황 및 데이터를 설명한다. 제4장에서는 문제 해결을 위한 예측 모델 개발 및 성능에 대해서 다룬다. 마지막으로 제5장에서는 본 연구의 결론을 제시한다.

2. 선행 연구

물류 분야에서는 효율성 증대를 위해 다양한 연구가 진행되어 왔으며, 특히 최근에는 머신러닝을 활용한 연구가 많이 진행되었다. Ahn et al.[2]에서는 휘발유 출하와 관련된 물류 문제를 해결하기 위해 GLM, Random Forest, Deep Learning 등 여러 머신러닝 기반 알고리즘을 앙상블(Ensemble)한 모델을 제안했으며, Jun and Song[5]은 항만 물동량 예측 문제에서 회귀분석과 인공신경망 모델의 성능을 적합성, 신뢰성, 강건성 측면에서 비교분석을 진행하였다. 이외에도 군집 분석을 활용하여 효율적 배송을 위한 유동적 물량에 대한 균등 배송 분배[4] 등도 찾아볼 수 있다. 하지만 본 연구에서 다루는 배송 트럭의 무게 예측 모델에 대한 연구는 찾아볼 수 없었다.

본 연구에서 활용한 XGBoost Regressor는 머신러닝 기

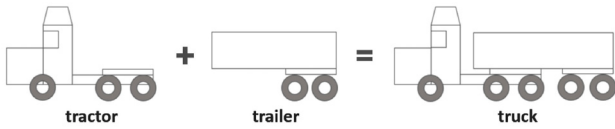
반 알고리즘으로 데이터가 희소한 경우에도 적용 가능하며 회귀, 분류, 랭킹에 뛰어난 성능을 보이는 앙상블 트리 모델이다. 간단하지만 데이터의 양이 많은 상황에서도 빠른 계산 속도 및 우수한 예측 정확도를 가지며 높은 수준의 범용성을 가지고 있으며, 행과 열로 표시가 가능한 정형데이터에서 특히 높은 성능을 보여준다[10]. XGBoost는 매우 다양한 연구에서 활용되고 있으며, Yoon et al.[14]에서는 산업 기술 수준에 영향을 미치는 요인 파악을 위해 활용하였고, Han and Kim[3]에서는 반도체 공정에서 가상 계측을 위한 예측 모델로 활용하였다. 이외에도 XGBoost를 활용한 다양한 분야의 연구들을 찾아볼 수 있다.

회귀모델을 학습할 때 종속 변수와 관련이 없는 변수를 제거하는 것은 효율적이고 정확한 성능의 모델을 구축하기 위해 필요하다. 많은 연구에서 피쳐 수를 줄이기 위해 다양한 방법을 사용하고 있는데, 주로 중요도가 높은 설명 변수를 선택하거나 요인 분석을 활용하여 새로운 변수를 추출한다. 본 연구는 Permutation Feature Importance(PFI)를 활용하여 중요도가 높은 변수를 추출한다. PFI는 특정한 피쳐를 사용하지 않았을 때, 모델의 성능이 감소되는 정도를 보고 특성의 중요도를 파악한다. 특히 변수의 수가 많은 비선형 모델에서 유용하며, 계산이 빠르고 일관된 변수의 중요도를 측정할 수 있는 특징을 갖아 다양한 분야에서 활용되고 있다[7]. 예를 들어, Lee[8]에서는 중소기업의 매출액에 영향을 미치는 예측 변수를 확인 시 PFI를 활용하였으며, Kang et al.[6]에서는 생애주기별 고혈압의 위험 인자를 파악하기 위해 PFI를 활용하였다[9].

3. 문제 상황

미국 캘리포니아 주에 한 맥주 회사는 Fairfield와 LA 두 곳의 양조장에서 맥주를 생산한다. 생산된 제품은 패키징하여 제3자 물류를 활용하여 미국 전역에 배송하고 있다. 제3자 물류 회사는 다수의 물류 회사들로 구성되어 있으며 다양한 적재가능용량을 가진 트럭들이 배송을 위해 할당된다. 이러한 과정에서 맥주 회사는 주문 수량 및 제3자 물류에서 제공한 배송 트럭의 적재 가능 용량을 고려하여 미리 제품의 패키징 및 트레일러 형성을 함으로써 배송 과정의 효율성을 꾀하고 있다.

<Figure 1>과 같이 배송 트럭은 주로 트랙터와 트레일러로 구성된다. 미국 웨스트버지니아 주 도로교통법에 따르면 트랙터는 주로 다른 차량을 견인하기 위해 설계되고 사용되는 모든 자동차를 의미한다. 트레일러는 동력이 있거나 없는 모든 운송수단으로, 사람이나 재산을 끌거나 자동차에 견인되도록 설계되었다. 따라서 제품을 트레일러에 적재하고 트랙터로 견인하여 제품 운송이 이루어진다.



<Figure 1> Configuration of Delivery Truck

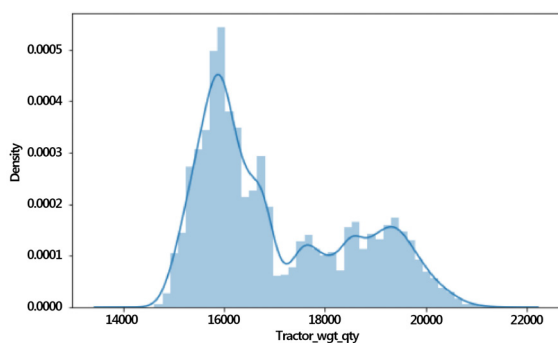
또한, 미국 국토부의 Compilation of Existing State Truck Size and Weight Limit Laws에서는 도로 안전 상의 이유로 트럭의 무게에 제한을 두고 있다[13]. 이에 따라 캘리포니아 주의 고속도로 위 트럭 크기 및 총 무게는 최대 80,000lbs이다. 즉 트랙터와 트레일러, 적재된 맥주의 총 무게는 80,000lbs를 넘어서는 안 되며, 따라서 맥주 회사에서는 제3자 물류 회사에서 제공하는 트랙터의 무게를 고려하여 미리 패키징할 맥주의 양을 결정하고 트레일러를 준비해야 한다. 너무 많은 양의 맥주를 미리 패키징하게 되면 배송트럭의 고속도로 총 무게 제한을 어기기 때문에 다시 패키징 작업을 해야 하며, 너무 적은 양의 맥주를 미리 패키징할 경우에는 배송의 효율성이 감소하거나 상황에 따라 재작업이 필요할 수도 있다.

하지만 맥주 회사는 향후에 어떤 배송 트럭이 할당될지 알 수 없는 상황에 있다. 즉, 각 배송에 대하여 어떤 물류 회사가 할당되는지는 알 수 있지만, 할당되는 배송 트럭의 무게는 미리 알 수 없는 상황이다.

따라서 본 연구에서는 맥주회사의 데이터(2013년 12월 ~ 2014년 9월, 총 49,414건의 배송 데이터)를 활용하여 제3자 물류 회사명 및 기타 정보를 바탕으로 제공될 트랙터 무게를 예측하는 모델을 개발한다.

맥주회사에서 관리되고 있는 주요 데이터는 운송 시작 날짜(Week), 선적번호(Shipnum), 양조장(Brewery), 트랙터 무게(Tractor_wgt_qty), 운송요금(Pay_amount), 목적지까지의 거리(Miles_from_brewery), 목적지(City/State), 제3자물류회사명(Carrier) 등이다. 이 중 트랙터 무게는 본 연구에서 예측하고자 하는 변수이다.

배송트럭을 한번이라도 제공한 제3자 물류 회사는 총



<Figure 2> Weight Distribution of Tractors

126군데이며, <Figure 2>와 같이 제공된 트랙터의 무게는 최소 12,000lbs에서 최대 40,200lbs의 범위로 평균적으로 19,936lbs이다. 또한 적재한 맥주 무게는 최소 214lbs에서 최대 52,482lbs의 범위를 가진다. 트랙터와 트레일러, 적재된 맥주 무게를 모두 합하여 최소 28,629lbs에서 최대 79,921lbs의 값을 보이며 80,000lbs 이하로 유지한 것을 확인할 수 있었다.

4. 예측 모델

4.1 데이터 분석

전체 49,414건의 데이터 중 7,106건의 데이터는 운송요금, 운송거리, 목적지에 대한 값이 존재하지 않았다. 해당 데이터를 살펴본 결과, 이들은 모두 수출에 해당하는 배송건으로 별도로 관리하는 과정에서 정보가 제대로 입력되지 않아 누락된 것으로 추측되었으며, 본 연구에서는 미국 내에서 배송되는 배송건으로 제한하여 예측 모델을 개발하였다.

또한, 제공된 배송트럭이 트레일러 일체형 모델인 경우에는 트레일러 무게가 0으로 표시되었다. 하지만 이러한 트럭들은 본 연구의 궁극적인 목적인 사전 패키징 및 트레일러 준비에서 벗어나므로 이러한 데이터 또한 제거하였다.

따라서 이러한 데이터를 제거하고 총 46개의 물류회사 및 총 41,127건의 배송 데이터를 활용하여 예측 모델을 개발하였다.

4.2 일반적 예측 변수

본 연구에서는 맥주회사에서 기존에 관리되고 있는 주요 데이터 외에도 예측 성능을 향상하기 위해 다양한 요인을 추가하거나 새로운 파생 변수를 생성하여 주요 예측 요인에 대해 분석하였다.

출발지 역할을 하는 Brewery 변수는 FFD 또는 LA 두 가지로 분류된다. 이때 두 장소 모두를 출발지로 삼는 트럭회사와 FFD 또는 LA 하나의 출발지를 갖는 회사가 존재한다. 따라서 출발지의 특성에 따라 할당되는 물류회사 및 배송트럭이 영향을 받을 수 있으므로, 두 출발지를 모두 담당하는 경우 0, FFD만을 담당하면 1, LA만 담당하면 2로 다중 분류하여 Brew_type라는 변수를 생성하였다.

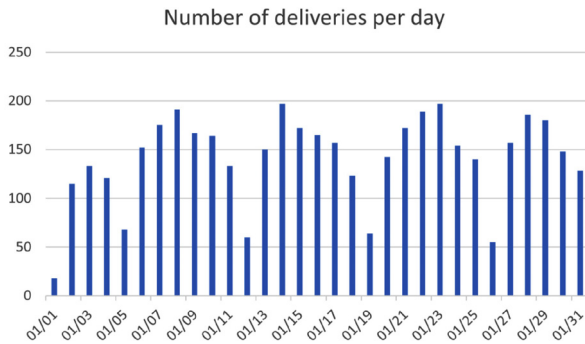
또한 목적지의 인구 특성, 즉 목적지의 맥주 배송 수요에 의해 배송트럭의 할당이 영향을 받을 수 있으므로, 미국 통계청에서 제공하는 인구수 추정 자료를 활용하여 2019년 주별 인구데이터 또한 예측 변수로 활용하였다 [12]. 또한 이를 위해 맥주의 소비층을 고려하여 20살 이상

의 인구 데이터만을 포함하였다.

많은 경우 날씨 또한 주요 예측변수로서 활용되므로 본 연구에서는 평균 온도, 총 강수량, 평균 풍속을 예측 변수로 포함하였다. 비, 폭풍우, 안개 유무 같은 변수의 경우에는 해당 기간에 대해 발생 횟수가 많지 않아 이상치로 취급될 수 있기 때문에 제외하였으며, 출발지인 FFD와 LA의 기상 데이터를 사용하였다. 날씨의 경우 Tutiempo의 데이터를 활용하였다[11].

또한 공시되는 캘리포니아 주의 주간 유가 데이터를 또한 예측 변수로 활용하였다. 이는 매주 월요일, 캘리포니아 주의 약 1,000개의 휘발유 판매점에서의 휘발유 판매 가격의 평균 가격이다.

<Figure 3>은 일별 배송 건수를 나타낸 것이다. 일요일 전후 또는 1월 1일 전후에 배송 건수가 적은 것을 알 수 있으며, 이러한 패턴 또한 배송트럭의 할당에 영향을 끼칠 수 있으므로 예측 변수에 포함하였다. 즉, 일요일 및 공휴일 여부 및 요일을 나타내는 변수를 예측 변수로 활용하였다.



<Figure 3> Number of Deliveries Per Day

4.3 운영 데이터에 기반한 예측 변수

본 연구에서는 예측의 정확도를 향상시키기 위해서 일반적인 예측 변수들 외에도 최근 운영 데이터에 기반한 파생 변수를 생성하여 예측 변수로 활용하였다. 즉, 각 물류 회사별로 운영하는 배송 트럭은 한정적이며 최근에 이미 할당하여 배송 중인 배송 트럭은 새로운 배송에 할당될 수 없으므로, 이를 설명하는 파생 변수를 생성하였다. 이를 위해 예측 시점 기준으로 최근 3일, 5일, 또는 7일 동안에 특정 무게 구간의 배송 트럭이 사용된 배송 건수를 바탕으로 해당 무게 구간의 배송 트럭이 가용할 확률을 계산하였다.

이를 위해 우선 제3자 물류 회사별 트랙터 무게에 대한 사분위수를 구하여 회사별 트랙터를 Q1, Q2, Q3, Q4로 분류하였다. 이때 Q1은 무게 기준으로 해당 물류 회사의 0백분위 이상 25백분위 이하의 트랙터 집합, Q2는 25백분위

초과 50백분위 이하의 트랙터 집합, Q3는 50백분위 초과 75백분위 이하의 트랙터 집합, Q4는 75백분위 이상 100백분위 이하의 트랙터 집합을 의미한다. 그 결과 46개의 물류 회사에 대해 각 4개의 그룹이 생성되었다.

최근 d 일을 기준으로 물류회사 c 가 보유 중인 무게그룹 Q_q 에 해당하는 트랙터의 개수를 $K_{c,q,d}$ 라고 정의하고, 본 연구에서는 보유 중인 데이터를 바탕으로 아래와 같은 식을 통해 $K_{c,q,d}$ 값을 추정하였다.

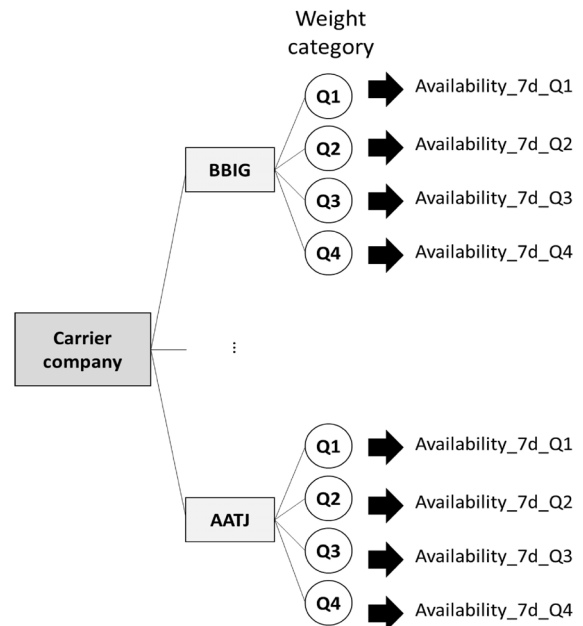
$$K_{c,q,d} = \max_{t=d, \dots, T} \sum_{x=t-d+1}^t n_{c,q,x}$$

이때, $n_{c,q,t}$ 는 t 일자의 배송에 할당된 물류회사 c 의 그룹 Q_q 에 트랙터의 대수이며, T 는 모델을 학습하기 위해 사용된 데이터의 마지막 일자를 의미한다. 본 연구에서는 d 를 위해 3, 5, 7일을 사용하였다.

이를 바탕으로 t 일에 추정하는 최근 d 일을 기준으로 물류회사 c 의 무게그룹 Q_q 에 해당하는 트랙터의 가용 확률 $P_{c,q,d}(t)$ 에 대한 식은 아래와 같다.

$$P_{c,q,d}(t) = \frac{K_{c,q,d} - \sum_{x=t-d}^{t-1} n_{c,q,x}}{K_{c,q,d}}$$

<Figure 4>는 본 연구에서 활용한 이러한 운영 데이터에 기반한 예측 변수(operational statistics parameter)를 표현한다.



<Figure 4> Estimated Availability for Each Carrier Company Based on 7 Days

4.4 예측 변수의 중요도 순위

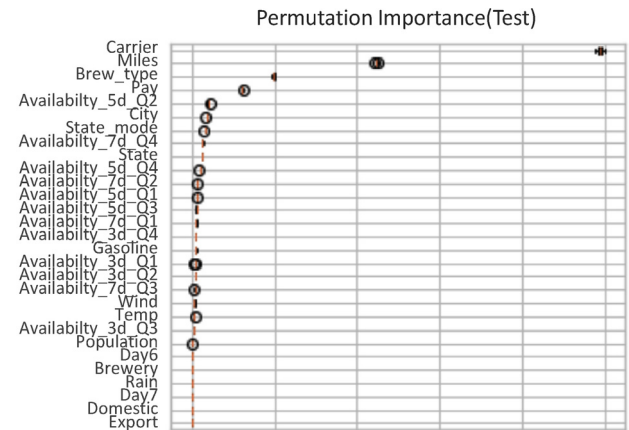
학습 모델은 XGBoost Regressor을 사용하였으며, 본 연구에서 활용된 피쳐(feature)는 <Table 1>에 정리되어 있다. 또한, 트랙터 무게 예측 시 변수의 범위와 단위의 영향을 줄이기 위해 급여, 운송 거리 변수에 대하여 StandardScaler를 활용하여 정규화를 실시하였으며, 제3자 물류 회사명, 양조장 등과 같은 범주형 데이터는 원핫(One-Hot) 인코딩을 적용하였다.

<Table 1> Feature Information for the Model

Feature name	Description	Type
Carrier	Name of 3PL company	Object
Brewery	Location of brewery	Object
Pay	Payment for delivery	Float
Miles	Distance of delivery	Float
State	State of destination	Object
City	City of destination	Object
Week	Date of delivery start	Object
Shipnum	Shipment number	Object
Export	Export or not	Binary
Trlr_wgt_qty	Weight of trailer	Float
Gasoline	Average price of gasoline at delivery date	Float
Temp	Average temperature at delivery date	Float
Rain	Average precipitation at delivery date	Float
Wind	Average wind speed at delivery date	Float
Day	Day of delivery	Object
Holi/Sun	Weekends/holidays or not	Binary
Brew_type	Type of brewery	Int
Population	Population of destination state	Int
State_mode	Mode destination of 3PL company	Object
Availability_n d_Qq	Estimated availability for weigh category Qq based on n days (q = 1,2,3,4, d = 3,5,7)	Float

피쳐 수에 비해 데이터의 양이 적어 permutation feature importance를 활용하여 피쳐별 중요도 순위를 확인하고 이를 통해 모델에 사용할 피쳐를 선정하였다. Permutation feature importance는 XGBoost와 같이 내부 원리를 정확히 파악할 수 없는 모델, 즉 black-box모델에 대하여 특정 변수를 쓰지 않았을 때 이것이 성능 하락에 얼마만큼의 영향을 미치는지를 확인하여 해당 변수의 중요도를 파악하는 방법이다. <Figure 5>는 이러한 permutation feature importance를 활용하여 각각의 피쳐에 대한 중요도를 시각화한 결과이다. 이에 따르면 트랙터 무게를 예측하는데 제3

자 물류 회사가 가장 큰 중요도를 보였으며, 다음으로는 운송거리, 양조장 타입, 운송요금, 트랙터의 가용 확률 $P_{c,2.5}(t)$ 등이 큰 중요도를 보였다. 또한 중요도 순위 상위권에 운영 데이터를 기반한 예측 변수가 다수 포함되었음을 보아 본 연구에서 개발한 트랙터의 가용 확률이 예측 성능 개선을 위해 효과적인 것을 기대할 수 있다.



<Figure 5> Permutation Feature Importance

4.5 예측 모델

피쳐별 중요도를 바탕으로 예측 모델에서 활용할 피쳐를 선정하기 위해, 먼저 제3자 물류 회사만을 활용한 모델을 개발한 후, 중요도가 높은 순서대로 피쳐들을 추가하며 모델의 성능을 분석하였다. 예를 들어, Model 1은 가장 중요도가 높았던 피쳐만을 사용하였고, Model 2는 중요도가 높은 두 개의 피쳐를, Model 3은 중요도가 높은 세개의 피쳐를 사용하였다. 모든 Model들은 동일하게 XGBoost Regressor를 사용하였다. 이러한 결과는 <Table 2>에 나타나 있다.

모델 성능 평가를 위해서는 MAPE(mean absolute percent error)를 활용하였으며, 통계적 분석을 위해 K-fold 교차검증을 30번 실행하여 얻은 MAPE 값들을 바탕으로 가설 검정한 결과는 <Table 2>와 같다. <Table 2>의 p-value는 Model 1과 해당 모델과의 단측 가설검정 결과이다.

Model 1은 중요도가 가장 높은 제3자 물류 회사만을 피쳐로 활용한 모델로서 MAPE 값은 2.918%로 이 자체만으로도 상당히 정확한 예측을 보여주었다. 이는 각 물류회사별로 보유하고 있는 트랙들의 무게 분포가 어느 정도 일정하고, 물류회사별로 그 분포에 차이가 있기 때문이라고 예상된다. 예측의 정확도 향상을 위하여 추가적인 피쳐들을 활용하여 모델을 테스트하였으며, Model 2는 중요도가 가장 중요한 두 피쳐(물류회사와 운송거리)를 활용하였고, Model 3은 물류회사, 운송거리, 양조장 타입, Model 4는

물류회사, 운송거리, 양조장 타입, 운송요금, Model 5는 물류회사, 운송거리, 양조장 타입, 운송요금, 트랙터의 가용 확률 $P_{c,2,5}(t)$ 를 피처로 활용하였다. 이러한 과정에서 대체적으로 개선된 성능을 보여줬으나, Model 5가 가장 좋은 MAPE인 2.785%의 성능을 보여줌과 동시에 처음으로 유의수준 3%에서도 통계적으로 유의미한 개선을 보여주었다. 이를 통해 본 연구에서 개발한 트럭 가용 확률이 예측 인자로 효과적임을 확인할 수 있으며, 특히 5일을 기준으로 한 확률 계산이 유의미함을 알 수 있다. 중요도가 높은 6개 또는 9개의 피처를 활용한 모델들(Model 6 및 Model 9) 또한 Model 1 대비 개선된 결과를 보여주었으나, 추가적인 검증 결과 Model 5 대비 개선된 결과를 보여준 않았다(유의수준 5%). Model 11 이후의 결과들 또한 Model 5 대비 성능 향상을 보이지 못했다. 이렇듯 사용된 피처수의 증가가 무조건적인 성능 향상을 야기하지 않았으며, 이는 다중공산성 문제, 불필요한 피처들로 인한 주요 피처의 가중치 적용 악영향 등에서 기인한 것으로 예상된다.

<Table 2> Performance of the Models

Models	MAPE		p -value
	Avg.	Var.	
Model 1	2.918	0.079	-
Model 2	2.819*	0.072	0.084
Model 3	2.819*	0.072	0.083
Model 4	2.820*	0.071	0.085
Model 5	2.785***	0.051	0.024
Model 6	2.785***	0.052	0.025
Model 7	2.917	0.068	0.494
Model 8	2.790**	0.062	0.034
Model 9	2.785***	0.056	0.026
Model 10	2.913	0.067	0.471

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.03$.

5. 결론

본 연구에서는 실제 제3자 물류를 활용하여 제품을 배송하고 있는 맥주 회사에 대해 배송 트럭의 무게를 예측하였다. 배송 트럭의 무게를 정확히 예측한다면 회사는 미리 맥주의 패키징 및 적재 준비를 할 수 있으며, 이를 통해 배송시간의 절약 및 효율성 증대를 기대할 수 있다. 이러한 무게 예측을 위해 해당 기업에서 관리되고 있는 데이터 외에도 일반적으로 예측을 위해 활용되는 날씨, 유가 등을 활용하였으며, 특히 과거의 운영 데이터를 바탕으로 각 물류회사 및 무게그룹별 배송 트럭의 가용 확률을 추정하는

예측 변수를 개발하여 적용하였다. 예측 모델을 위해서 XGBoost Regressor가 사용되었으며, 모델에 활용할 예측 변수 선정에 위해 Permutation Feature Importance 알고리즘을 활용하여 각각의 중요도를 확인하였다. 가장 중요도가 높았던 제3자 물류회사를 시작으로 중요도가 높은 변수들을 순서대로 추가하며 모델을 개발한 결과, 물류회사, 운송거리, 양조장 타입, 운송요금, 트랙터의 가용 확률을 피처로 활용한 모델이 가장 좋은 MAPE인 2.785%의 성능을 보여주었다. 특히, 운영 데이터를 기반으로 개발한 트랙터의 가용 확률을 추가하였을 때 통계적으로 유의미한 정확도 향상을 확인할 수 있다.

더 다양한 사례들에 대한 본 연구의 예측 변수(특히, 운영 데이터를 바탕으로 한 트럭 집단별 가용 확률 추정 변수) 활용 및 성능 평가는 본 연구의 향후 연구 방향이 될 수 있을 것이다.

Acknowledgement

This work was supported by Yonsei Business Research Institute and National Research Foundation of Korea (NRF) grant funded by the Korea government(MSIT) (No. 2022R1C1C101173111).

References

- [1] Aggarwal, V. and Parameswaran, L., Effect of overweight trucks on fatigue damage of a bridge. *Advances in Structural Engineering*, Springer, New Delhi, 2015, pp. 2483-2491.
- [2] Ahn, S., Jang, D., and Park S. Comparing Gasoline Shipment Prediction Model Performance of Oil Reservoir using Machine Learning, *The Journal of Korean Institute of Information Technology*, 2020, Vol. 18, No. 11, pp. 17-23.
- [3] Han, J. and Kim, H., XGBoost Based Prediction Model for Virtual Metrology in Semiconductor Manufacturing Process, *Proceedings of the Korea Information Processing Society Conference*, 2022, Vol. 29, No. 1, pp. 477-480.
- [4] Jeon, H. and Lim, H., A Study on Optimization for Delivery Destination Clustering using Unsupervised Learning, *Proceedings of the Korea Information Processing Society Conference*, 2022, Vol. 29, No. 1, pp. 483-486.
- [5] Jun, C.Y. and Song, J.M., A Study on Application of the Port Cargo Throughput Forecast by the Neural Network Model, *The Journal of Shipping and Logistics*, 2007, Vol. 53, pp. 65-82.

- [6] Kang, S., Kim, S., and Ryu, M.H., Analysis of Hypertension Risk Factors by Life Cycle Based on Machine Learning, *Journal of the Korea Industrial Information Systems Research*, 2022, Vol. 27, No. 5, pp. 73-82.
- [7] Loh, W.Y., Fifty years of classification and regression trees, *International Statistical Review*, 2014, Vol. 82, No. 3, pp. 329-348.
- [8] Lee, G., Exploring the predictive variables of major economic-related statistical indicators on SME sales using machine learning: Focusing on small and medium-sized businesses in print-related shopping malls, *The Journal of Korean Association of Computer Education*, 2022, Vol. 25, No. 3, pp. 79-89.
- [9] Ramirez, S.G., Hales, R.C., Williams, G.P. and Jones, N.L., Extending SC-PDSI-PM with neural network regression using GLDAS data and Permutation Feature Importance, *Environmental Modelling and Software*, 2022, Vol. 157, p. 105475.
- [10] Shwartz-Ziv, R. and Armon, A., Tabular data: Deep learning is not all you need. *Information Fusion*, 2022, Vol. 81, pp.84-90.
- [11] ToTiempo.net, <https://en.tutiempo.net>.
- [12] United State Census Homepage, <https://www.census.gov>.
- [13] USDOT Regulatory Relief Fact Sheet, 2014, <https://www.nemaweb.org/index.php/docman/nema-forumsmeetings/2014-nema-forums/2014-annual-forum/183-usdot-regulatory-relief-fact-sheet-june-2014/>.
- [14] Yoon, Y.A., Jung, J.H., Lim, J.H. Chang, T., and Kim Y. S., A study on Data Preprocessing for Developing Remaining Useful Life Predictions based on Stochastic Degradation Models Using Air Craft Engine Data, *Journal of the Society of Korea Industrial and Systems Engineering*, 2020, Vol. 43, No. 2, pp. 48-55.

ORCIDYu-jin Lee | <https://orcid.org/0000-0002-7584-9397>Kyeong Min Choi | <https://orcid.org/0000-0002-1981-8422>Song-eun Kim | <https://orcid.org/0000-0002-3688-1853>Kyungsu Park | <https://orcid.org/0000-0002-5386-5222>Seung Hwan Jung | <https://orcid.org/0000-0002-3044-3879>