

머신러닝 기법을 이용한 총생산시간 예측 연구

남은재* · 김광수*

*한국교통대학교 산업경영공학과

A Study on Total Production Time Prediction Using Machine Learning Techniques

Eun-Jae Nam* · Kwang-Soo Kim*

*Department of Industrial Engineering, Korea National University of Transportation

Abstract

The entire industry is increasing the use of big data analysis using artificial intelligence technology due to the Fourth Industrial Revolution. The value of big data is increasing, and the same is true of the production technology. However, small and medium-sized manufacturers with small size are difficult to use for work due to lack of data management ability, and it is difficult to enter smart factories.

Therefore, to help small and medium-sized manufacturing companies use big data, we will predict the gross production time through machine learning. In previous studies, machine learning was conducted as a time and quantity factor for production, and the excellence of the ExtraTree Algorithm was confirmed by predicting gross product time.

In this study, the worker's proficiency factors were added to the time and quantity factors necessary for production, and the prediction rate of LightGBM Algorithm knowing was the highest. The results of the study will help to enhance the company's competitiveness and enhance the competitiveness of the company by identifying the possibility of data utilization of the MES system and supporting systematic production schedule management.

Keywords : Production Time Prediction, Machine Learning, Algorithm

1. 서론

전 산업계는 4차산업혁명으로 인공지능 기술을 이용한 빅데이터 분석의 활용도가 점차 높아지고 있다. 클라우드 슈밥(Klaus Schwab)의 다보스 세계경제포럼 이후 빅데이터(Big data)의 가치는 확장되고 있으며, 진입이 다소 늦은 생산 제조기술 분야의 활용도와 관심 또한 점차 확대되고 있다[13].

본 연구에서는 중소기업의 스마트공장 기초단계에서 구축된 MES(Manufacturing Execution System) 시스템의 데이터를 활용하여 인공지능 모델링 분석으로 생산 시간 예측을 구현하고자 한다.

MES는 자동차, 전자, 반도체 산업 등 제조업 부문에 실

시간 생산계획에 따른 진행현황 파악을 위해 폭넓게 사용되고 있다. 중소기업의 경우 데이터 관리 운영이 어려워 자료수집 용도로 활용되고 있으나 중기부에서는 계속해서 스마트공장의 고도화 추진으로 방향을 잡고 기초단계 지원을 제한하고 있다.

이에 기업 자체적으로 MES 시스템의 축적된 생산 실적 빅데이터를 활용하여 Machine Learning Algorithms을 통해 총생산시간을 예측함으로써 체계적인 생산 일정 관리가 될 수 있다. 이로써 중소기업 스마트공장 고도화의 추진 의지를 강화하고 기업의 경쟁력 제고에 도움이 되고자 한다.

†Corresponding Author : Kwang-Soo Kim, Department of Industrial & Management Engineering, Korea National University of Transportation, 50, Daehak-ro, Daesowon-myeon, Chungju-si, Chungcheongbuk-do, Republic of Korea, E-mail: kskim@ut.ac.kr
Received May 15, 2023; Revision June 13, 2023; Accepted June 26, 2023

2. 이론적 배경

주문생산방식의 제조기업에서 복잡한 환경하에 제품의 제작에 필요한 총생산시간을 예측하는 것은 매우 의미 있는 작업이다. 이를 위해 총생산시간 산정에 영향을 주는 요인을 파악하는 것이 선행되어야 한다.

본 연구에서 사용하게 될 주요 요인에 대해 설명하고 대량의 데이터를 빠르게 처리할 수 있는 Machine Learning Algorithms을 활용하여 기업의 경영성과 제고에 효과적인 총생산시간 예측을 추정하고자 한다.

2.1 생산시간 및 생산 일정

테일러 (Frederick W. Taylor)에 의해 과업관리의 한 수단으로 오늘날 널리 활용되고 있는 시간연구(Time Study)는 소정의 작업을 수행하는데 소요되는 허용시간, 즉 표준시간의 산정기법과 이론적 근거를 제시하며, 이러한 기법을 적용하는 연구 분야이다[6]. 제조의 총시간에는 불필요한 많은 요소를 포함한다. 제조의 가치를 높이기 위해 비생산적 요소를 적극적으로 제거하고 최적의 작업 방법에서 이루어진 표준작업시간을 설정하는 것이 시간연구의 최종 목적이다[10].

2.2 총 생산시간의 개념

총생산시간이란 제품의 생산이 시작되는 시점부터 완료되는 시점까지 걸리는 시간의 합으로 총생산시간을 구하는 식은 아래와 같다[2].

$$TPT = \sum_{i=0}^n ST_i + \sum_{i=0}^{n-1} DT_i$$

(Number of processes : $i = 0, 1, 2, \dots, n$)

산출식에서 TPT(Total Production Time)는 총생산시간을 의미하며 각각의 공정에서 걸리는 표준시간(St: Standard Time)의 합과 공정 전·후 사이의 대기시간(DT: Delay Time)의 합을 더한 것이다.

2.2.1 표준시간

표준시간이란 일반적으로 숙련된 작업자가 규정된 방법을 사용하여 지정된 작업을 수행하는 데 필요한 시간으로 정의하고[14] 표준시간을 계산하는 식은 아래와 같다[15].

$$\text{Standard time} = \text{Normal Time} + \text{Allowable Time}$$

$$\text{Standard time} = \text{Normal Time} \times (1 + \text{Allowance Rate})$$

표준시간은 준비시간과 주 작업 시간으로 구성되며 준비 시간 및 단위당 시간은 각각 정미 시간과 여유시간으로 이루어진다. 여기에서 정미 시간은 작업의 기본내용을 구성하는 규칙적·주기적으로 반복되는 작업부분에 요구되는 시간으로 수정이 필요한 경우 수행도 평가(Performance Rating)를 수행한다[7].

2.3 생산성 및 숙련도 선행연구

글로벌 공급망 확보 경쟁과 기업 간의 다양한 갈등 속에서 중소 제조기업의 인력난은 매우 심각하다. 현장에서는 인력을 구하지 못해 원활한 공장가동이 힘든 상황이다. 구인난은 저출산과 고령화로 인해 생산가능인구의 감소도 큰 원인이 되고 있다. 이에 기업에서는 적은 인원으로 높은 생산성을 올리기 위한 다양한 노력을 시도하고 있으며 숙련도와 생산역량에 관한 연구에 높은 관심을 두게 되었다.

4M은 공정을 구성하는 장비(Machine), 표준작업방법(Method), 작업자(Man), 자재/부품(Material)을 의미하는 것으로 제조업에서 생산성을 향상하고, 불량률을 관리하는 데 중요한 역할을 한다. 제조업 생산관리의 효율적 관리를 위해 생산 현장의 시시각각 변화하는 생산활동 자원(4M)을 실시간으로 자료화하고 분석하기 위한 연구결과와 작업자의 역량에 따라 불량수량의 차이가 크게 발생한다는 것을 확인하고 작업자에 대한 체계적인 교육과 훈련이 필요한 것으로 조사하였다[9].

업무숙련도와 관련한 연구에서 이용주 (1999)는 대기행렬 시스템에서 업무숙련도가 생산성 향상에 영향이 미침을 확인하였고 제조업에서 자동화 기술의 발전과 함께 기술적 숙련이 기업의 경쟁력 유지에 결정적인 역할을 한다는 것을 확인하였다[11]. 또한, 점차 기술이 고도화되어가고 있는 산업현장에서 기계, 설비 등의 발전과 정보화로 시스템 측면의 개선과 함께 시스템을 운영하고 제품을 생산하는 중심에 있는 작업자의 역량 강화를 통한 낭비 요소 제거, 생산성 향상이 매우 중요하고 그 효과가 크다는 것을 확인했다[1].

선행연구의 결과를 통해 제조현장에서 작업자의 기술력 숙련도가 생산성에 큰 영향을 미친다는 것을 인식하고 있음에도 계속해서 변화하는 작업자의 역량을 체계적으로 관리하고 효과를 분석하는 것에는 어려움이 있다. 이에 본 연구에서는 생산 정보화 시스템의 데이터를 사용하여 작업자의 시간당 생산 수량을 산출하고 총생산시간을 예측하는 인공지능 모델링 학습에 설명변수로 사용하고자 한다.

2.4 인공지능을 활용한 생산시간 연구

제조업은 국가 경제성장의 견인차 역할을 하고 있으나 최근 경기침체 및 노동 인력의 감소, 인건비 상승 등의 이유로 운영에 어려움을 겪고 있다. 이러한 어려움 극복을 위해 인공지능을 활용한 생산연구가 활발히 진행되고 있다. 박찬범 (2017)은 뿌리 산업 생산공정 예측모델 개발 연구에 인공지능을 활용하였고 Machine Learning 을 이용하여 일별 관중 수 예측과 비정상적 관중 수를 예측하는 연구가 진행되었다[8].

이처럼 다양한 분야의 인공지능을 활용한 연구를 통해 제조업의 경쟁력 제고와 위기를 극복하기 위한 요소로 제조현장의 데이터를 기반으로 의사 결정할 수 있는 공장 운영시스템을 갖추는 것이 중요하다[5]. 산업 전반에 인공지능 모델을 적용하려는 시도가 활발히 진행되고 있는 것은 지속적인 기술의 발달로 인공지능 모델링의 정확도가 높아지고 있기 때문이다.

주문생산방식 기업의 기계학습 기반 총생산시간 예측 연구로 생산 현장으로부터 수주 자료와 제품의 특징, 생산 실적의 자료를 수집하여 선형 알고리즘과 비선형 알고리즘을 적용한 회귀분석을 통한 총생산시간을 추정하는 연구에 OLS, GLM Gamma, Extra Tree, Random Forst의 다양한 기법을 활용하였으며 Extra Tree알고리즘을 적용한 모델이 최고의 성능을 보이는 것으로 조사되었다 [2].

본 연구에서는 자동차부품 제조기업의 총생산시간 예측 연구로 선행연구에서 우수성능을 확인한 Extra Tree 알고리즘 외 다양한 알고리즘을 통한 모델링 학습 결과를 통해 더 향상된 알고리즘을 찾고 총생산시간 예측에 영향을 미치는 요소를 조사하였다.

2.4.1 특성검토

RandomForest는 데이터 세트의 다양한 하위 샘플에 대한 여러 분류 결정 트리를 맞추는 메타 추정기로 평균화를 사용하여 예측정확도를 개선하고 과적합을 제한한다 [19].

2.4.2 알고리즘

예측기법은 과거 데이터를 이용하여 미래를 예측하는 회귀분석이 주로 사용되는데 회귀분석은 변수 간의 관계를 설명하고 독립변수에 영향을 받는 종속변수를 예측하는 목적으로 사용된다[4]. IBM은 예측분석에 대해 새로운 개념이 아닌 의사결정 트리와 로지스틱 회귀 분석을 사

용하여 기업이 데이터를 분류하고 예측하는 것이라고 정의하였다. 점차 예측분석 범위가 넓어지면서 Machine Learning과 Deep Learning의 실용성이 적어 투자를 주저했지만, 점차 직관적이고 새로운 예측기법과 하이브리드 클라우드 구축 모델을 통해 분석이 쉬워졌기 때문에 기업이 경쟁력을 확보하기 위해 도입을 주장하고 있다[16].

본 연구에서는 총생산시간을 추정하는 Machine Learning Algorithms 기법으로 <Table 1>과 같이 ExtraTree, GradientBoosting, Hist_Gradient [19], XGBRegressor [17], LGBM[18] 알고리즘을 사용하였다.

<Table 1> Machine Learning Algorithm

Algorithm	Explanation
ExtraTree	Implement a meta estimator that fits multiple random decision trees to various subsets of the dataset, improves prediction accuracy using averaging, and controls overfitting.
Gradient Boosting	It performs extremely well on extensive datasets and is one of the most widely used techniques for structured classification and regression prediction modeling problems. The main issue is that the model training speed is slow.
Hist_Gradient	To address the issues of gradient boosting, it discretizes continuous input variables into hundreds of distinct values and trains trees that are added to the ensemble, thus increasing the speed.
XGBRegressor	It is a model that evolved GBM (Gradient Boosting Machine) using an ensemble of Decision Trees. By altering the structure of the learning model for performance optimization, it improved the system and enhanced the execution speed.
LightGBM	It is a gradient boosting framework that uses tree-based learning algorithms and includes parameters to limit tree depth, reducing overfitting. It demonstrates fast training speed, high efficiency, and accuracy.

2.4.3 성능검증

회귀분석의 예측모델 성능평가 지표는 결정계수(R^2), 평균제곱오차(MSE), 평균절대오차(MAE)를 많이 사용한다. 본 논문에서 본 연구에서는 MSE, RMSE, MAE, R^2 및 예측모델의 정밀도와 재현율의 조화평균값인 Score를 예측모델의 성능평가로 사용하였다.

3. 연구방법

3.1 연구문제

제조업의 총생산시간 예측 모델링 연구에서는 표준시간과 대기시간을 반영한 총생산시간의 관점으로 접근하여 비선형 모델의 Extra Tree 모형의 성능이 가장 우수한 것으로 조사되었다[3]. 본 연구에서는 표준시간과 비가동시간에 작업자별로 생산 수량의 평균값을 반영하여 총생산시간의 관리범위를 확대하였고 Extra Tree 모형 이외 더 다양한 알고리즘을 사용하였다는 점에서 차별점을 가진다.

가) 기존의 표준시간과 대기시간을 반영한 총생산시간의 관점에서 표준시간, 실제 시간, 개인별 시간, 비가동시간을 반영하여 총생산시간의 관리범위를 확대하였다.

나) ExtraTree, GradientBoosting, HistGradient Boosting, XGBRegressor, LightGBM 알고리즘을 적용하여 성능을 비교한 결과 LightGBM알고리즘의 예측율이 가장 높은 것으로 확인되었다.

4. 데이터 분석

본 연구에서는 자동차부품을 제조하는 A 기업의 실제 사례로 MES 시스템에서 제품의 생산기록을 추출하여 분석 데이터로 사용하였다. 자료를 제공한 회사는 2015년부터 MES 시스템을 도입하여 생산관리의 정보화를 추진하였으나 BOM 등 기초 데이터 구축의 어려움과 작업현장에서 시스템 활용의 어려움으로 실패를 거듭하다 2019년 이후로 데이터 수집의 안정화가 이루어졌다. 자료수집의 안정화로 축적된 데이터의 분석을 통한 계획관리를 구현하고자 제품의 생산실적기록 정보를 추출한 후 기계학습 알고리즘에 맞게 전처리를 진행하였다.

4.1 MES시스템 데이터 추출

본 연구에 사용된 자료의 기본 구성은 <Table 2>와 같이 번호, 작업일자, 작업공정, 작업그룹, 차종, 제품코드, 제품명, 롯트번호, 작업자 이름, 투입인원수, 주간/야간, 총생산수량, 양품수량, 불량수량, 표준UPH, 실제UPH, 총작업시간, 비가동시간, 계획비가동시간, 작업시작일자, 작업시작시간, 작업종료일자, 작업종료시간, 생산실적번호,

발주번호 기계학습 알고리즘에 사용하기 위해 필요항목을 선정하고 개인별 숙련도가 생산성에 미치는 영향을 분석하기 위해 개인의 해당 기간의 제품별 시간당 생산수량의 중앙값을 개인별UPH로 추가하였다.

<Table 2> Table Structure

No	Field	Type	Explanation
1	No	Integer	Number
2	date	datetime	Date
3	job	Varchar	work process
4	job_grp	Varchar	work group
5	Item_grp	Varchar	car type
6	item	Varchar	item code
7	item_name	Varchar	item name
8	lot_no	Varchar	lot
9	name	Varchar	name
10	people	Integer	people
11	day_gub	Varchar	day/night
12	t_qty	Integer	total qty
13	o_qty	Integer	Good Quantity
14	e_qty	Integer	Poor Quantity
15	s_UPH	Float	standard UPH
16	r_UPH	Float	actual UPH
17	p_UPH	Float	individual UPH
18	t_time	Integer	total time
19	e-time	Integer	downtime
20	p_time	Integer	planned downtime
21	s_date	Date	work start date
22	s_time	Integer	work start time
23	e_date	Date	work end date
24	e_time2	Integer	work end time
25	prod_no	Varchar	production number
26	order_no	Varchar	order number

4.2 특징검토

A기업으로부터 제공받은 데이터는 ‘Object’타입에 대해 더미변수처리와 표준화로 전처리 하고 설명변수 선정을 위해 RandomForstRegressor로 Drop-Column Importance 방식으로 중요도(Feature importance)를 분석하였다.

분석결과 작업공정(전조), 작업공정(단조), 작업공정(OBJ조립), 작업공정(B/H전조), 작업공정(END조립)이 예측값에 미치는 영향이 작고 계획비가동시간(p_time), 작업그룹(job_grp), 총생산수량(t_qty), 개인별UPH(p_UPH), 양품수량(o_qty) 순으로 영향을 주는 것으로 조사되었다.

<Table 3> Table Structure of production result

Field	Feature_Importance	Field	Feature_Importance
job_grp	0.1280	p_time	0.5632
t_qty	0.0885	job_B/H_1	0.0056
o_qty	0.0436	job_B/H_2	0.0013
e_qty	0.0076	job_END	0.0052
s_UPH	0.0138	job_IBJ	0.0096
r_UPH_y	0.0223	job_OBJ	0.0008
p_UPH_y	0.0487	job_cold	0.0002
e_time	0.0413	job_flat	0.0002

중요도(Feature importance) 분석결과를 바탕으로 작업공정에 대한 7가지 항목과 불량수량(e-qty)을 제외하고 훈련모델의 성능을 높이기 위하여 RandomizedSearchCV를 통해 하이퍼파라미터튜닝과 모델훈련(fitting)을 진행하여 예측값에 대한 모델 검증을 진행하였다.

5. 연구결과

MES 시스템의 생산데이터를 수집하여 중요도 분석을 실시하고 인공지능 모델링 분석에 사용된 설명변수는 <Table 4>와 같다. 최종 선정된 설명변수로 인공지능 모델 훈련을 통해 자동차 부품 제조기업의 총생산시간을 예측한 결과 LightGBM 알고리즘이 가장 우수한 것으로 확인되었다.

<Table 4> Table Structure of production result

Field	Type	Explanation
job	Varchar	work process
job_grp	Varchar	work group
Item_grp	Varchar	car type
item	Varchar	item code
item_name	Varchar	item name
lot_no	Varchar	lot
name	Varchar	name
people	Integer	people
day_gub	Varchar	day/night
t_qty	Integer	total qty
o_qty	Integer	Good Quantity
e_qty	Integer	Poor Quantity
s_UPH	Float	standard UPH
r_UPH	Float	actual UPH
p_UPH	Float	individual UPH
t_time	Integer	total time

Field	Type	Explanation
e-time	Integer	total time
p_time	Integer	planned downtime
s_date	Date	work start date
s_time	Integer	work start time
e_date	Date	work end date
e_time2	Integer	work end time
prod_no	Varchar	production number
order_no	Varchar	order number

5.1 모델 검증

모델 검증결과 MAE 14.27, MSE 896.01, RMSE 29.93, R2 Score 0.82의 값을 보인 LightGBM Model이 선택되었다.

<Table 5> Model Validation Results

Model	MAE	MSE	RMSE	R2
ExtraTrees	15.2573	995.4943	31.5514	0.8023
GradientBoosting	15.6823	985.3406	31.3901	0.8043
HistGradientBoosting	14.3582	899.6729	29.9945	0.8213
XGBRegressor	15.7017	983.527	31.3612	0.8046
LGBM	14.2772	896.0114	29.9334	0.8220

5.2 연구의 의의

총생산시간 추정과 관련한 연구는 이전부터 꾸준히 진행되고 있다. 인공지능 기술이 발전하면서 생산관리 분야에도 연구가 활발히 진행되고 있으며 점차 더 우수한 성능의 회귀모형 알고리즘이 개발되면서 제조업 분야에 많은 활용을 기대하고 있다.

본 연구에서는 자동차부품 제조기업의 MES 시스템 데이터를 확보하여 연구를 진행함으로써 인공지능 기술의 적용이 어려운 제조기업에 활용 가능성을 보여주었고 작업자의 숙련도를 총생산시간 추정에 활용함으로써 생산성과 숙련도의 연관성을 확인하였다는 점에서 큰 의미가 있다.

5.3 향후 연구 방향

인공지능 모델에 대한 전 세계적인 관심과 발전으로 Machine Learning을 통한 예측의 활용 수요는 점점 높아질 것이다. 그러나 모델링 학습에 필요한 기초데이터의 종류와 정확도에 따라 전혀 다른 결과를 가져올 수 있다. 제조현장의 MES 정보는 일반적으로 작업자가 수동으로

결과를 입력하는 형태이며, 세팅 시간, 고장시간, 대기시간 등 설명변수의 좀 더 다양한 세분화가 필요하다.

본 연구를 통해 MES 시스템의 제조현장 데이터로 인공지능 모델링 활용의 가능성을 확인하였고 작업자의 숙련도와 생산성과의 연관성을 확인하였다. 향후 좀 더 다양한 데이터 확보를 통해 인공지능 모델링을 통한 예측성능을 향상하고자 한다. 또한, 데이터 분석연구에 그치지 않고 제조기업에 실용화할 수 있는 시스템 구축에 관한 연구를 진행함으로써 기업의 경영성과 제고에 이바지하고자 한다.

6. References

- [1] C. U. Bae, M. H. Mun, J. S. Kim(2016), A study on productivity improvement methods through NCS-Based worker competency diagnosis. Korean Society of Management Science, pp. 1089-1094.
- [2] D. M. Park(2021), "A study on how to apply machine learning to estimate total production time in individual production methods." Doctoral dissertation, Donga University Graduate School.
- [3] D. M. Park, H. R. Choi, B. K. Park(2021), "Machine learning-based total production time prediction method for custom-made companies." Intelligence Information Research, 27(1):177-190.
- [4] G. H. Lee(2022), "Correlation analysis between factors to improve smart factory production." Doctoral dissertation, Korea Polytechnic University.
- [5] H. G. Kim(2019), "An empirical study on the intention of continuous use and conversion of smart factories." Doctoral dissertation, Pusan National University Graduate School of International Studies.
- [6] H. Whang(2005). Work management theory. Yeongji Cultural History.
- [7] H. Y. Kim(1996). "Standard time calculation method by real-time data collection."
- [8] J. H. Yoo(2019), "Study on prediction of attendance using machine learning." Journal of IKEEE, 23(4):1243-1249. doi: 10.7471/IKEEE.2019.23.4.1243
- [9] J. S. Kim, W. S. Cho(2015), "Analysis of 4M data in small and medium-sized manufacturing processes for the introduction of big data." Korean Journal of Data Information Science, 26(5):1117-1128.
- [10] J. S. Lim, M. C. Lim, M. C. Um(2004). Operations management. Hyeongseol.
- [11] Y. J. Noh, M. R. Kim(2020), "An empirical study on the factors influencing technical skill development of production workers." Vocational Competency Development Research, 23(3):139-168.
- [12] B. Bahu, L. Bironneau, V. Hovelaque(2019), "Comprehension du DDMRP et de son adoption : Premiers elements empiriques." Logistique & Management, 27(1):20-32.
- [13] K. Schwab(2017), The Fourth Industrial Revolution: Currency. Crown Business, New York, p. 4.
- [14] Z. Kjell, M. Harold(2001), Maynard's industrial engineering handbook (5th ed.). McGraw-Hill Education.
- [15] P. G. Milkell(2006), Work systems and the methods, measurement, and management of work. Pearson.
- [16] IBM(2022, November 7). Decision tree. Retrieved from <https://www.ibm.com/downloads/cas/GQP5QPXZ>
- [17] Kaggle XGBM(2022, October 27). XGBM. Retrieved from <https://www.kaggle.com>
- [18] LightGBM(2022, November 2). LGBM algorithm. Retrieved from <https://lightgbm.readthedocs.io/en/v3.3.2/>
- [19] Sklearn(2022, November 2). randomforestregressor. Retrieved from <https://scikit-learn.org/stable/modules/classes.html#>

저자 소개



남 은 재

중앙대학교 글로벌인적자원개발대학원에서 석사 취득. 현재 한국교통대학교 산업경영공학과 대학원 재학 중.

관심분야 : 인적자원개발, 생산관리, 빅데이터 분야



김 광 수

한국교통대학교 산업경영 안전 공학부에 재직 중이며 대한 안전경영과학회 이사와 한국품질경영학회 부회장을 역임 하였으며, 관심분야: 품질경영, HRD 분야