

# 주문생산 기업을 위한 기계학습 기반 총생산시간 예측 기법

박도명

동아대학교 경영정보학과  
(home21cc@gmail.com)

최형림

동아대학교 경영정보학과  
(hrchoi@dau.ac.kr)

박병권

동아대학교 경영정보학과  
(bpark@dau.ac.kr)

4차 산업혁명 기술의 발전으로 사람이 처리하지 못하는 부분을 기계학습 등 인공지능 기법을 활용하여 개선해 보려는 노력이 확대되고 있다. 주문형 생산 기업에서도 주문에 대한 총생산시간을 예측하여 납기 지연 등의 기업 리스크를 줄이고자 하나 주문마다 총생산시간이 모두 달라 이를 예측하는데, 어려움을 겪고 있다. 주문 처리량 증대, 주문 총비용 절감을 위해 효율성이 가장 낮은 영역을 찾아 그 영역을 강화하는 TOC(Theory of constraints) 이론이 개발되었으나 총생산시간 예측은 제시하지 못하였다. 주문생산은 고객의 다양한 요구로 인해 주문마다 그 특성이 모두 다르므로 개별적인 주문의 총생산시간을 사후에 측정할 수는 있으나 사전 예측을 하기는 어렵다. 기존 주문의 이미 측정된 총생산시간도 모두 달라 표준 시간으로 활용할 수 없는 한계성이 있다. 이에 따라 경험이 많은 관리자는 시스템의 이용보다는 감에 의존하고 있고, 경험이 부족한 관리자는 간단한 관리지표(예, 원재료가 파이프이면 총생산시간 60일, 철판이면 총생산시간 90일 등)를 사용하고 있다. 불완전한 감이나 지표를 기초로 하여 작업 지시를 너무 빨리하면 정체가 발생하여 생산성이 저하되고, 너무 늦게 하면 긴급 처리로 인해 생산비용이 증가하거나 납기를 지키지 못하는 경우가 발생한다. 납기를 지키지 못하면 지체상금을 배상해야 하거나 영업, 수급 등의 부문에 악영향을 미친다. 본 연구에서는 이러한 문제를 해결하기 위하여 주문생산시스템을 운영하는 기업의 신규 주문 총생산시간을 추정하는 기계학습 모델을 찾고자 한다. 기계학습에 활용된 자료는 수주, 생산, 공정 실적을 사용한다. 그리고 총생산시간의 추정에 가장 적합한 알고리즘으로 OLS, GLM Gamma, Extra Trees, Random Forest 알고리즘 등을 비교 분석하고 그 결과를 제시하고자 한다.

**주제어** : 주문생산, 총생산시간, 기계학습

논문접수일 : 2021년 1월 26일    논문수정일 : 2021년 3월 7일    게재확정일 : 2021년 3월 18일  
원고유형 : 일반논문    교신저자 : 최형림

## 1. 서론

제 2차 산업혁명으로 산업이 발달하면서 기업은 생산성과 효율성을 높이기 위해 노력해왔다. 생산성과 효율성의 측정을 위해서는 개선 전과 개선 후를 비교되는 기준이 필요하다. 프레드릭 테일러(F. W Taylor)는 과학적 관리법에서 시간 연구와 동작연구를 통해 단위 작업에 대한 표준 작업량과 표준 시간을 설정하려고 노력하였다

(Taylor, F. W, 2016). 프랭크와 릴리언 길브레스(F and L, Gilbreth) 부부 또한 동작연구를 통해 비능률적이고 잘못된 동작으로부터 생긴 낭비를 없애는 노력을 하였다(Wikipedia, 2020).

고객의 요구가 다양해지면서 다품종 소량 생산에 필요한 생산시간을 측정하고 활용하는 연구도 진행 중이다(Lee, 2012; Choi and Kwon, 1998). 측정된 생산시간을 준비, 가공, 정리 등 요소작업으로 분류하여 측정의 횟수를 줄이는

PTS(Predetermined Time Standards)법의 활용이 연구되고 있다(Park and Choi, 1987).

주문형 생산시스템의 연구로 생산 일정 계획을 수립하기 위해 이종 혹은 동종 병렬 기계 군에는 유전자 알고리즘을 적용하고 단일 기계에는 최소 유효시간(Minimum slack)의 규칙을 적용한 특허가 있으나 공정의 시간과 작업 순서의 효율적 배정에 목표를 두고 있다(Kim, D. W and J. Y. Lee, K. H. Cho, 2012). 또 다른 연구로 ERP 패키지에서 제공하는 전진 계산법과 후진 계산법을 적용하여 작업시간, 대기시간, 이동시간을 일자별로 작업장의 부하로 쌓아가며 계산하는 방식이 있으나 제품별 공정별 작업시간이 사전에 정의되어있어야 사용 가능한 방법이다(Lee and Lee, 1999).

선행 연구된 주문형 생산시스템의 시간 연구는 공정 단위의 작업시간이 사전에 정의되어 있다는 조건을 만족한 상태에서 일정 관리나 작업 배정을 하는 연구는 이루어지고 있으나 총생산 시간을 예측하는 연구는 찾을 수 없었다. 그 이유는 주문생산을 하는 기업은 주문 시점에 제품의 규격과 제조 요구사항이 확정된다. 제품이 과거에 생산된 적은 있으나 고객의 요구에 따라 규격과 제작 공정의 요구사항이 달라지기 때문이다. 본 연구에서 사용한 자료를 바탕으로 분석한 결과 연간 기준 60% 정도 새로운 제품의 주문이 일어나고 있다. 새로운 제품의 생산에 필요한 총 시간은 실제로 생산이 수행된 후 사후에 측정하는 것은 가능하지만 주문 시점에서 사전에 정확히 예측하는 것은 매우 어렵다.

새로운 제품은 새로운 생산시간의 측정을 필요하게 하고, 과거 생산시간의 재활용률을 낮게 만든다. 즉 생산시간의 재활용률이 낮다는 것은 생산시간을 측정하는 담당자가 불만을 품도록

한다.

이런 이유로 경험이 많은 관리자는 시스템의 이용보다는 감에 의존하고 있고, 경험이 부족한 관리자는 간단한 관리지표(예, 원재료로 파이프를 사용하는 제품은 생산시간이 60일 소요되고, 철판을 사용하는 제품은 생산시간이 90일 소요)를 만들어 사용하고 있다.

불완전한 생산시간을 기초로 작업 지시를 너무 빨리하면 공정의 정체 현상이 발생하고 생산성은 저하된다. 너무 늦게 하면 특급 작업으로 생산 비용이 증가하고 납기를 지키지 못하게 된다.

납기를 지키지 못하면 지체상금을 배상해야 하거나, 영업직과 고객의 사이가 나빠지게 된다. 그뿐만 아니라 수금의 지연으로 기업의 재무 안정이 나빠지게 된다. 이러한 문제를 해결하기 위해 제품의 생산이 시작되는 시점부터 완료되는 시점까지 걸리는 시간의 합인 총생산시간의 추정은 꼭 필요하다. 따라서 본 연구에서는 주문생산 기업의 복잡한 환경하에서 제품의 제작에 필요한 총생산시간을 예측하기 위해 적합한 기계 학습 알고리즘을 찾고 총생산시간을 추정하고자 한다.

본 논문의 구성은 다음과 같다. 제2장에서 분석에 사용된 알고리즘의 간단한 소개와 알고리즘 선정 이유를 설명하였다. 제3장에서는 데이터의 설명, 설명변수의 선택과정에 대하여 설명한다. 제4장에서는 알고리즘별 결과를 요약하고 종합적으로 분석하여 설명한다. 마지막으로 제5장에서 결론을 맺는다.

## 2. 선행 연구

### 2.1 분석 알고리즘

다양한 기계학습 방법이 있으나 선형 회귀 분석 알고리즘으로 OLS(Ordinary Least Squares), GLM(Generalized Linear Model)가 있다. 비 선형 회귀 분석 알고리즘은 Ada Boosting, Extra Trees, Gradient Boosting, Random Forest, SVR(Support Vector Regression), Stacking, Voting 모델이 있다. 딥 러닝 모델로 Multi-Layer Preceptorn 모델이 있다. 본 연구에서는 OLS, GLM Gamma, Random Forest, Extra Trees 모델을 선택하여 비교 분석하였다.

#### 2.2.1 OLS(Ordinary Least Squares)

총생산시간은 각각의 작업 요소 시간과 작업의 수량을 곱하여 계산되는 선형 모델로 OLS(Ordinary Least Squares)모델 채택하였다(Scikit-learn developers, 2007~2020; Perter and John, 1989).

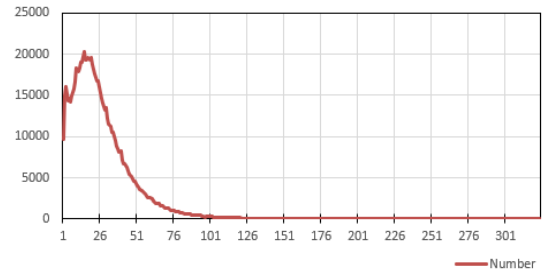
#### 2.2.2 GLM(Generalized Linear Model) Gamma

학습용 데이터의 종속 변수인 실 생산시간이 같은 횟수를 누적하면 <Figure 1>과 같은 감마분포( $k = 3, \theta = 2$ )의 형태를 보여 GLM(Generalized Linear Model)을 채택하였다(Scikit-learn developers, 2007~2020; Perter and John, 1989).

#### 2.2.3 Random Forest

다수의 의사결정트리를 만들고 트리를 분류 집계하여 최종적으로 분류하는 모델로 431가지

의 다양한 Feature를 사용하는 본 논문에 적합한 모델로 선택하였다(Park, 2018).



<Figure 1> Number of Production times

#### 2.2.4 Extra Trees

랜덤 포레스트에서 파생된 모델로 많은 수의 의사 결정 트리를 무작위로 분할 한 다음 그중에서 최상의 분할을 선택하는 알고리즘으로 편향은 늘어나지만, 분산은 낮아지게 되는 특성을 가진다(Park, 2018).

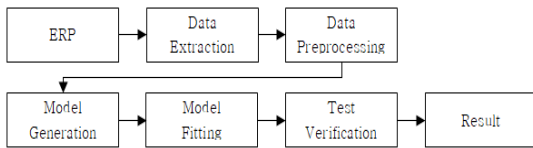
복잡한 설명변수를 가진 본 연구는 다양한 트리의 분할이 필요하다. 트리 분할의 무작위성이 좋은 모델로 Random Forest와 Extra Trees가 있으며, 항 간염 웹 타이트 연구에서 157개의 유효한 설명 인자를 채택하는데 효과를 보인 연구가 있다(Mishra and Sehgal, 2017).

본 연구에서 사용한 데이터가 복잡하고 다양한 설명변수를 가지고 있어 선형(Linear)보다 비선형(Non-Linear)이 더 적합할 것으로 예상된다. 예측되어야 하는 종속 변수의 값이 다양해야 하므로, Random Forest와 Extra Trees 알고리즘을 선택하였으나 무작위성이 큰 Extra Trees 알고리즘의 성능이 우수할 것으로 예상하였다.

### 3. 데이터와 분석 방법

#### 3.1 데이터

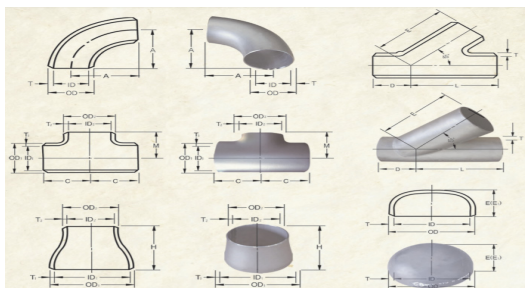
본 연구에서 데이터 분석 절차는 ERP 시스템으로부터 과거에 생성된 수주 데이터, 제품의 생산 기록, 제품의 특성 자료를 추출한다. 추출된 데이터를 기계학습 알고리즘에 맞게 전 처리한다. 분석에 필요한 모델을 생성하고 모델 피팅을 실시한다. 테스트 데이터를 사용하여 검증 절차를 거치고 결과를 분석한다. 상세한 내용은 <Figure 2>에서 확인할 수 있다.



<Figure 2> Data analysis procedure

본 연구에서 사용한 데이터는 40년간 관 연결구를 제조해 온 회사로 전 세계 시장의 30%의 점유율을 가지고 있는 A 기업이다.

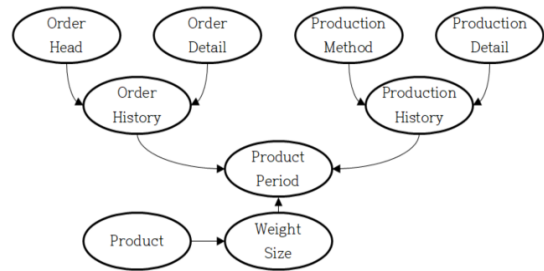
주 생산 품목으로 엘보우(Elbow), 티(Tee), 레듀서(Reducer), 레트랄(Lateral), 캡(Cap) 이 있다. 제품의 형상은 <Figure 3>에서 확인할 수 있다.



<Figure 3> Main Products

연구 대상 데이터는 납기 일자가 2008년 01월 01일 ~ 2019년 12월 31일의 범위에 포함된, 자체 생산한 엘보(Elbow)를 대상으로 하였다.

고객의 수주 데이터, 제품의 생산 기록, 제품의 특성 자료를 수집하여 사용하였으며, <Figure 4>의 과정에 따라 데이터를 추출하였다.



<Figure 4> Data extraction process

또한, 정제된 데이터를 얻기 위해, 생산 중에 제조 공정이 분할된 공정의 생산 완료 시간은 최댓값을 적용하였다. 제조 기록의 관리번호가 11자리가 아닌 것은 삭제하였다. 엘보우 제품과 관련된 238개의 공법을 정리하여 분석에 사용하였다. 공정분석의 편리성을 위하여 다양한 이름으로 사용 중인 공정을 하나의 공정으로 변환하였다. 제품의 제조는 공법에 종속되거나 고객의 요구로 수행하는 공정에 있어 작업을 수행한 기록이 있는 공정은 1, 없는 공정에 대해서는 0으로 변환하여 공정 391개를 분석에 사용하였다. 총생산시간을 계산하기 위하여 제조 기록의 시작 시각과 완료 시간의 차이를 구하고, 총생산시간에서 휴일(토, 일요일)을 차감하여 실 생산시간을 구하였다. 공정 간의 작업 부하를 반영하기 위하여 수주 시점의 공정중량, 납기 시점의 공정중량을 계산하여 추가하였다. <Table 1>은 1차 전 처리된 데이터로 Tuple(188,659), Feature(431)가 선

〈Table 1〉 Row data set

Number	Data Type	Feature Name	Number of groups	Feature Description
1	Categorical	RECEIPT_NO	174,883	Order Number
2	Numerical	SEQ_NO		Order sequence
3	Categorical	G_GUBN	1	Product classification code
4	Categorical	PROCESS_CODE	238	Production method code
5	Categorical	RECEIPT_DATE	-	Order date
6	Categorical	DELI_DATE	-	Delivery date
7	Numeric	MARGIN	-	Orders - Delivery period
8	Numeric	QTY	-	Order quantity
9	Numeric	LOT_QTY	-	Input quantity
10	Numeric	E_QTY	-	Completed quantity
11	Categorical	CURRENCY	7	Foreign currency type
12	Numeric	EXCHANGE_RATE	-	exchange rate
13	Numeric	UNIT_PRICE	-	Unit Price
14	Numeric	AMOUNT_HAN	-	the amount in Korean won
15	Numeric	DEG	-	Angular measure
16	Categorical	G_GILD	2	Plating code
17	Categorical	G_WELD	28	Welding code
18	Categorical	G_TYPE	55	Bevel code
19	Categorical	G_GLASS	20	Material rating code
20	Categorical	G_ITEM1	1	Product group code
21	Categorical	G_ITEM2	11	product code
22	Categorical	G_MAT1	47	Material group code
23	Categorical	G_MAT2	189	Material code
24	Categorical	G_SCH1, 2	1155, 29	Thickness 1, 2 code
25	Categorical	G_SIZE1, 2, 3	159, 2, 1	Size1, 2, 3 code
26	Numeric	CSCH1_MM	-	Thickness 1 value
27	Numeric	CSCH2_MM	-	Thickness 2 value
28	Numeric	CSIZE1_MM	-	Size1 value
29	Numeric	CSIZE2_MM	-	Size2 value
30	Numeric	CSIZE3_MM	-	Size3 value
31	Numeric	CWEIGHT	-	Product weight
32	Categorical	MAT_GROUP	6	Raw material group code
33	Numeric	RWEIGHT	-	Product weight based on order date
34	Numeric	DWEIGHT	-	Product weight based on delivery date
35	Categorical	PO_NO	156,395	Manufacturing management number
36	Categorical	C312	3,746	Date of completion of the cutting process
37	Categorical	C391	3,094	Date of completion of production
38	Numeric	TOTDAY	-	Total working time
39	Numeric	HOLIDAY	-	a Holiday time
40	Numeric	C001~C999	391	Process execution status
41	Numeric	EFFDAY	-	Working time

〈Table 2〉 Process code description

Process Code	Description	Process Code	Description
C031	1st pickling	C423	Final hot forming
C033	1st shorting	C454	Surface grinding
C050	2nd: Hot	C457	Surface scratch
C064	2nd shorting	C561	NDE(MT)
C120	2nd shorting	C562	NDE(MT,UT)
C124	Orthodontic part grinding	C563	NDE(PT)
C155	Final cold forming	C564	NDE(RT)
C212	Rolling	C566	PMI TEST
C278	Bead removal idea	C707	BALL calibration
C303	Molding	C712	Surface scratches
C306	Molding	C738	Heat treatment (Q)
C333	Hot forming	C758	Re-shorting
C336	Heat treatment (N)	C764	Re-shorting
C340	Heat treatment (S.T)	C798	1st: Hot
C344	Welding	C831	Hot forming
C359	Heat treatment (T)	C838	Bead removal idea
C376	True circle correction	C846	Attach support
C383	Attaching the support	C918	Grinding
C389	Remove support	C946	1st shorting
C392	Inspection	C987	NDE(PT)-surface
C394	Mounting	C988	NDE(PT)-bevel
C395	Installation idea		

정되었다. 범주형 데이터는 수치형 데이터로 변환하고, 수치형 데이터는 0과 1 사이의 값으로 정규화(Normalization)하였다. 분석에 사용한 데이터는 90:10의 비율로 학습데이터와 검증을 위한 테스트 데이터로 분리하였다.

### 3.2 설명변수의 선택

선형 알고리즘의 설명변수를 선택하는 방법은 Status Model을 결괏값에서 계산된 P-Value > 0.05인 Feature를 순차적으로 제거하였다. 다중공선성 문제 해결을 위해 VIF > 10인 경우 Feature

를 제거하였다.

비선형 알고리즘의 설명변수를 선택하는 방법은 Ada Boost, Gradient Boosting, Random Forest Model에서 Feature importance 값이 모두 0인 Feature를 순차적으로 제거하였다.

선형 알고리즘을 적용한 모델은 1차에서 895개, 2차에서 66개, 3차에서 3개의 Feature를 제거하고 4차에 걸친 설명변수 선택 방법을 적용하고 P-Value와 VIF을 만족하는 const, Margin(여유 시간), LotQty(제조로트수량), EndQty(재고수량), AmountHan(원화금액), Sch2(두께2), Size1(사이즈 1), Weight(중량), Rweight(수주기준중량), Dweight

〈Table 3〉  $R^2$  value of times

Model	Linear $R^2$ value					Non-Linear $R^2$ value				
	1 STEP	2 STEP	3 STEP	4 STEP	MSE	1 STEP	2 STEP	3 STEP	4 STEP	MSE
OLS	0.3777	0.3615	0.3618	0.3584	188.45	0.4143	0.4137	0.4084	0.4083	173.80
GLM Gamma	0.3871	0.3685	0.3686	0.3592	197.61	0.4503	0.4496	0.4414	0.4410	182.29
ExtraTrees	0.5823	0.5826	0.5785	0.5702	126.24	0.7317	0.7306	0.7245	0.7232	81.29
RandomForest	0.2991	0.2991	0.2952	0.2967	206.58	0.3121	0.3121	0.3121	0.3121	202.05

(납기기준중량), C031, C033, C050, C064, C120, C124, C155, C212, C278, C303, C333, C336, C359, C376, C395, C423, C454, C457, C561, C562, C564, C712, C738, C758, C764, C798, C838, C846, C884, C918, C946, C987의 설명변수가 채택되었으며 공정 설명변수 내용은 <Table 2>에서 설명하였다.

비선형 알고리즘은 1차에서 917개, 2차에서 6개, 3차에서 5개의 Feature를 제거하고 4차에 걸친 설명변수 선택 방법을 적용하고 Feature importance 값을 만족하는 const, Pr\_EES05(엘로우 타원화 스테인리스), Pr\_EHN12(엘로우 열간 노말라이징), Pr\_EHQ08(엘로우 열간 퀘칭 열처리), Pr\_EHS11(엘로우 열간 스테인리스), Pr\_ETC01(엘로우 투심 카본), Pr\_ETC03(엘로우 투심 카본), Pr\_ETQ01(엘로우 투심 퀘칭 열처리), Pr\_ETS09(엘로우 투심 스테인리스), CU\_A(달러), We\_00, We\_0R, Ty\_00, Ty\_20, Ty\_BF, Ty\_BU, Cl\_0, Cl\_3, Cl\_A, I2\_1, M1\_70, M1\_71, M2\_01, M2\_02, M2\_04, M2\_10, M2\_17, M2\_48, Margin(여유시간), Qty(수주수량), LotQty(투입수량), EndQty(재고수량), ExchangeRate(환율), UnitPrice(단가), AmountHan(원화), Degree(각도), Sch1(두께1), Sch2(두께2), Size1(사이즈1), Weight(중량), Rweight(수주기준중량), Dweight(납기기준중량), C031, C033, C064, C120, C124,

C278, C303, C306, C333, C340, C344, C359, C376, C383, C389, C392, C394, C423, C457, C561, C563, C564, C566, C707, C712, C738, C758, C764, C798, C831, C838, C846, C918, C946, C987, C988의 설명변수를 채택하게 되었다. <Table 3>에서  $R^2$ 값과 MSE(Mean Squared Error)값을 확인할 수 있다.

## 4. 분석 결과

### 4.1 분석 결과 요약

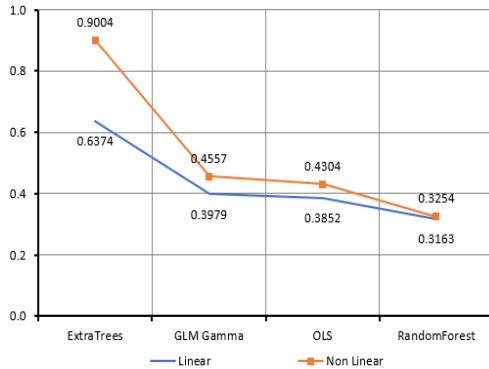
#### 4.1.1 선형, 비선형 알고리즘의 $R^2$ 값 비교

선형, 비선형 알고리즘의 비교 결과 전체적으로 비선형의 설명변수 선택기법을 적용했을 때  $R^2$ 값이 크게 나왔다. 이것은 비선형 모델의 적용이 총생산시간의 추정에 더 적합하다고 할 수 있다.

선형 알고리즘에서  $R^2$ 값은 Extra trees > GLM Gamma > OLS > Random Forest 알고리즘의 순서로 0.5702, 0.3592, 0.3584, 0.2967의 값을 보인다.

비선형 알고리즘에서  $R^2$ 값은 Extra Trees > GLM Gamma > OLS > Random Forest 알고리즘

의 순서로 0.7232, 0.4410, 0.4083, 0.3121의 값을 보인다. 상세한 내용은 <Table 3>과 <Figure 5>에서 확인할 수 있다.



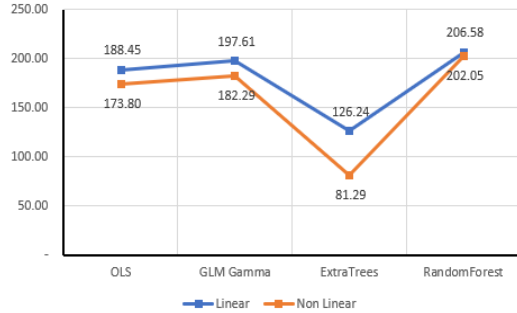
<Figure 5>  $R^2$  value of times

#### 4.1.2 MSE(Mean Squared Error)값 비교

선형, 비선형 알고리즘의 비교 결과 전체적으로 비선형의 설명변수 선택기법을 적용했을 때 MSE 값이 낮게 나왔다. 이것은 비선형 알고리즘의 적용이 총생산시간의 추정에 더 적합하다고 할 수 있다.

선형 알고리즘에서 MSE의 값은 Extra Trees < OLS < GLM Gamma < Random Forest 알고리즘의 순서로 126.24 < 188.45 < 197.61 < 206.58의 값을 보인다.

비선형 알고리즘에서 MSE의 값은 Extra Trees < OLS < GLM Gamma < Random Forest 알고리즘 순서로 81.29 < 173.80 < 182.29 < 202.05의 값을 보인다. 상세한 내용은 <Table 3>와 <Figure 6>에서 확인할 수 있다.



<Figure 6> MSE value of tiems

## 5. 결론

### 5.1 결과 요약

복잡한 설명변수를 갖는 주문생산 기업에서 총 생산시간의 추정은 Extra Trees 알고리즘이 적합하다. 분석에 사용된 설명변수 중 채택된 Pr\_EES05(엘보우 타원화 스테인리스), Pr\_EHN12(엘보우 연간 노말라이징), Pr\_EHQ08(엘보우 연간 켄칭 열처리), Pr\_EHS11(엘보우 연간 스테인리스), Pr\_ETC01(엘보우 투심 카본), Pr\_ETC03(엘보우 투심 카본), Pr\_ETQ01(엘보우 투심 켄칭 열처리), Pr\_ETS09(엘보우 투심 스테인리스), CU\_A(달러), We\_00, We\_0R, Ty\_00, Ty\_20, Ty\_BF, Ty\_BU(제품타입), Cl\_0, Cl\_3, Cl\_A, I2\_1, M1\_70, M1\_71, M2\_01, M2\_02, M2\_04, M2\_10, M2\_17, M2\_48(재질)은 범주형 변수로부터 생성된 자료이다. C031, C033, C064, C120, C124, C278, C303, C306, C333, C340, C344, C359, C376, C383, C389, C392, C394, C423, C457, C561, C563, C564, C566, C707, C712, C738, C758, C764, C798, C831, C838, C846,



C918, C946, C987, C988은 1의 값을 갖는 특수한 형태의 숫자 형 설명변수다. 나머지 Margin(여유 시간), Qty(수주수량), LotQty(투입수량), EndQty(재고수량), ExchangeRate(환율), UnitPrice(단가), AmountHan(원화금액), Degree(각도), Sch1(두께 1), Sch2(두께2), Size1(사이즈1), Weight(중량), Rweight(수주기준중량), Dweight(납기기준중량)는 숫자 형 설명변수다.

총제조시간을 구한 추정 식은 위에서 나열한 27개의 범주 형 설명변수와 1의 값을 가지는 36개의 숫자 형 변수 그리고 14개의 숫자 형 변수로 구성된다. 특히 선형 알고리즘과 다르게 공법과 제품 타입, 재질이 총제조시간의 설명변수로 추가되어 영향을 미치는 것으로 확인되었다.

본 연구에서 다수의 범주형 데이터를 사용함으로 무작위성이 높아야 한다. 이는 Extra Trees의 특성과 부합 하여 좋은 결과가 만들어진다. Extra Trees 알고리즘을 적용하여 총생산시간을 추정한 결과  $R^2$  0.7232, MSE 81.29의 값을 가진다.

## 5.2 이론적 공헌

총생산시간 추정과 관련한 기존의 연구는 개별 공정의 시간 연구나 작업의 분배에 초점을 맞추고 있었으나 본 연구는 제품 제작의 전체적인 관점에서 총생산시간을 중요하게 생각하고 접근하였다. 그뿐만 아니라. 회귀분석이 가능한 OLS, GLM Gamma, Extra Trees, Random Forest 알고리즘을 적용하여 비교하였다는 점에서 기존 연구와 차별성을 가진다.

## 5.3 실무적 공헌

산업적인 측면에서도 기계학습을 활용하여 총

생산시간을 추정하게 함으로써 생산시간의 측정과 입력에 필요한 시간을 줄일 수 있다. 생산의 시작 시점을 알 수 있으므로, 근로자가 계획단계에서 더 많은 시간을 할애할 수 있다. 특히 업무의 부하로 놓치는 일이 발생하더라도 시작일과 현재일 비교를 통해서 진행되지 않은 업무를 파악하고 대응해 나갈 수 있다.

## 5.4 연구한계

본 연구에서 2가지의 연구 한계성을 가지고 있다. 첫 번째는 A 업체에서 제조하는 ELBOW 제품만을 반영한 것으로 같은 설비에 다른 제품을 같이 생산하는 경우 총제조시간에 영향을 줄 수 있다. 두 번째는 외주에서 반제품 또는 완제품 구매를 통해서 입고되는 제품을 제외하였다. 외주에서 입고되는 제품의 조달 시점은 내부에서 생산하는 총제조시간과 다른 시간이 될 수 있다. 그뿐만 아니라 조달 업체의 데이터를 추출할 수 없는 문제점을 가지고 있어 총제조시간을 추정할 수 없었다.

## 5.5 향후 연구 방향

본 연구 결과, Extra Trees 알고리즘의 오차가 가장 적음을 확인하였으므로 향후 실무에서 이를 이용하여 총생산시간을 추정하기 위해서는 다음이 필요하다. 첫째, 기계학습을 위해 다양한 형태의 자료를 변환하는 기법이 연구되어 있다. 그러나 원시 데이터를 기계학습에 맞는 데이터로 변환하는 작업은 아직도 기초적인 기술만 정리되어 있다. 기계학습에 필요한 데이터 형태를 분류하고 라이브러리를 작성한다면 많은 도움이 될 것으로 생각한다. 둘째, 총생산시간 예측을 자동화하기 위해서는 Extra Trees 알고리즘을 활

용한 과업관리 시스템 개발이 필요하다.

## 참고문헌(References)

- Aleksei, G., “Histogram-Based Algorithm for Building Gradient Boosting Ensembles of Piecewise Linear Decision Trees”, *International Conference on Analysis of Images, Social Networks and Texts*, (2019), 38~50
- Alexey, N., K. Alois, “Gradient boosting machines, a tutorial”, *US National Library of Medicine National Institutes of Health*, (2013)
- Cho, H. J., J. I. Park, “A Method for Construction Productivity index for Increasing Productivity”, *Korean Institute Of Industrial Engineers*, (2010), 832
- Choi, J. H., H. J. Kwon and J. Y. Woo and P. Park, “A case study of work improvement adapting standard time and line balancing on a manufacturing process”, *Journal of the Korea Institute of Plat Engineering*, Vol3 (1998), 172
- Davis, F. D., “Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology”, *MIS Quarterly*, Vol.13(1989), 319~342
- Dwight, D. E., “A speech to the National Defense Executive Reserve Conference in Washington, DC”, (1957)
- Goldratt, E., *The Goal second revised edition.*, Dongyangbooks, (2002)
- Epstein, M. J., M. J. Roy, “Improving sustainability performance: Specifying, implementing and measuring key principles”, *Journal of General Management*, Vol.29(2003), 15~31
- Friedman, M., *The social responsibility of business is to increase its profits*, The New York Times Magazine, September 13 (1970)
- Harris, D., *Improving Regressors using Boosting Techniques*, (1997)
- Kim, D. W., J. Y. Lee and K. H. Cho, “Algorithms for establishing production schedule plans for custom production plants and systems for establishing production schedule plans using them”, *Korea Intellectual Property Office*, (2012)
- Kim, G. H., “A Study on the Evaluation of MBO Operation and Effectiveness”, *The Korea Association for Policy Studies*, Vol13(2004), 52~54
- Kim, S. M., J. K. Ahn, “Verification of ERP Standard Time Using TOC Technique and Improvement of MES Routing Point”, *Journal of Society of Korea Industrial and System Engineering*, Vo141(2018), 30~31
- Kim, S. Y., H. Y. Soon and S. W. Choi, “Intelligent Digital Factory for Productivity Innovation: Remodeling of A Tractor Factory”, *Korea Productivity Association*, Vol23(2009), 6~7
- Kim, Y. Y., “Manufacturing Innovation and HPC(High Performance Computing) Utilization”, *Korea Technology Innovation Society*, Vol19 (2016), 250
- Lee, D. H., S. Y. Cho and H. W. Kim, “A Study on the Increase of Satisfaction and Productivity of Smart Work”, *Korea Society of Management Information Systems*, (2013), 298~299

- Lee, J., H. S. Kang, *Environmental management theory*, Hyungseol publishing house, 2003
- Lee, K. J., “Establishment of Standard Time for Stranding Process of a Cable Company in Small Quantity Batch Production System”, *Journal of the Society of Korea Industrial and System Engineering*, Vol35(2012), 100 ~ 101
- Lee, S. K., Y. H. Lee, “Production Scheduling employing ERP in the make-to-order manufacturing system”, *IE Interfaces*, Vol12 (1999), 427~429
- Maltz, E., A. K. Kohla, “Market Intelligence Dissemination Across Functional Boundaries”, *Journal of Marketing Research*, Vol33(1996), 47~61
- Mishra, G. J., D. P. Sehgal, “Quantitative Structure Activity Relationship study of the Anti-Hepatitis Peptides employing Random Forests and Extra-trees regressors”, *Bioinformantion*, Vol13(2017),
- Navver, j., S. Slater, “The Effect of a Market Orientation on Business Profitability,” *Journal of Marketing*, Vol54(1990), 20~35
- Park, H. S., *Hands-On Machine Learning with Scikit-Learn & TensorFlow*, Hanbit Media Inc, (2018)
- Park, J. M., “A Study on the Perception of Corporate Members on the Factors of Productivity Improvement”, *Korea Productivity Association*, Vol10(1995), 178~181
- Park, S. H., S. W. Choi, “Establishment of Standard Times Using STD Technique in Multi-product and One-unit Production System”, *Korean Association of Industrial Business Administration*, Vol10(1987), 28~30
- Perter, M., N. John, *Generalized Linear Models, Second Edition*, Chapman and Hall/CRC, (1989)
- Ryu, H. K., “OLAP and Decision Tree Analysis of Productivity Affected by Construction Duration Impact Factors”, *Journal of the Korea Institute of Building Construction*, Vol11 (2011), 106
- Scikit-learn developers, User Guide, 2007~2020, [https://scikit-learn.org/stable/modules/linear\\_model.html](https://scikit-learn.org/stable/modules/linear_model.html)
- Scikit-learn developers, User Guide, 2007~2020, <https://scikit-learn.org/stable/modules/ensemble.html>
- Taylor., F. W, *The Principles of Scientific Management*, ModiBooks, (2016)
- Wikipedia, (2020), [https://en.wikipedia.org/wiki/Time\\_and\\_motion\\_study](https://en.wikipedia.org/wiki/Time_and_motion_study)
- Woo. M. S., J. W. Yoo. “Impact of relative investment in SFA technology on new product success through market information acquisition process and perceived intelligence quality: Moderation Effect of Market Uncertainty and Technical Turbulence”, *Korea Productivity Association*, Vol32(2018), 174~176

Abstract

## **A Machine Learning-based Total Production Time Prediction Method for Customized-Manufacturing Companies**

Do-Myung Park\* · Byung-Kwon Park\*\* · HyungRim Choi\*\*\*

Due to the development of the fourth industrial revolution technology, efforts are being made to improve areas that humans cannot handle by utilizing artificial intelligence techniques such as machine learning. Although on-demand production companies also want to reduce corporate risks such as delays in delivery by predicting total production time for orders, they are having difficulty predicting this because the total production time is all different for each order. The Theory of Constraints (TOC) theory was developed to find the least efficient areas to increase order throughput and reduce order total cost, but failed to provide a forecast of total production time. Order production varies from order to order due to various customer needs, so the total production time of individual orders can be measured postmortem, but it is difficult to predict in advance. The total measured production time of existing orders is also different, which has limitations that cannot be used as standard time. As a result, experienced managers rely on persimmons rather than on the use of the system, while inexperienced managers use simple management indicators (e.g., 60 days total production time for raw materials, 90 days total production time for steel plates, etc.). Too fast work instructions based on imperfections or indicators cause congestion, which leads to productivity degradation, and too late leads to increased production costs or failure to meet delivery dates due to emergency processing. Failure to meet the deadline will result in compensation for delayed compensation or adversely affect business and collection sectors. In this study, to address these problems, an entity that operates an order production system seeks to find a machine learning model that estimates the total production time of new orders. It uses orders, production, and process performance for materials used for machine learning. We compared and analyzed OLS, GLM Gamma, Extra Trees, and Random

---

\* Department of MIS, Dong-A University

\*\* Department of MIS, Dong-A University

\*\*\* Corresponding Author: Hyung Rim Choi

Department of MIS, Dong-A University

2-1 Bumin-dong, Seo-gu Hoegi-ro, Busan 602-760, Korea

Tel: +82-51-200-7400, Fax: +82-51-200-5615, E-mail: hrchoi@dau.ac.kr

Forest algorithms as the best algorithms for estimating total production time and present the results.

**Key Words** : Order production, Total production time, Machine learning

Received : January 26, 2021 Revised : March 7, 2021 Accepted : March 18, 2021

Corresponding Author : HyungRim Choi

## 저 자 소개



**박도명**

동아대학교 경영정보학과에서 석사학위를 취득하였으며, 현재 동아대학교 경영정보학과 박사과정에 있다. 중견 기업의 전산실 이사를 역임하고 있으며, 관심 분야는 기계학습, 인공지능, C#, Database, 스마트제조 등이다.



**최형림**

KAIST 경영과학과에서 박사학위를 취득하였으며, 현재 동아대학교 경영정보학과 교수로 재직중이다. 동아대학교 스마트물류연구센터와 산학연연구단지조성사업단의 단장을 역임하고 있으며, 관심 분야는 인공지능, 빅데이터, IoT, 해운항만물류 등이다.



**박병권**

KAIST 전산학과에서 박사학위를 취득하였으며, 현재 동아대학교 경영정보학과 교수로 재직중이다. 동아대학교 4차산업혁명혁신선도대학사업단의 단장을 역임하고 있으며, 관심 분야는 빅데이터분석, IoT, 클라우드컴퓨팅, 스마트헬스케어, 스마트제조 등이다.