

스마트공장 수준 진단 데이터 분석을 통한 업종별 평가영역별 제언

이민주¹ · 이지수¹ · 최지원¹ · 장태우^{1*} · 배재현²

¹경기대학교 산업경영공학과 / ²한국생산성본부

Improvement Suggestions by Industry based on Assessment Data Analysis of Smart Factories

Min-ju Lee¹ · Ji-su Lee¹ · Ji-won Choi¹ · Tai-woo Chang¹ · Jae-Hyun Bae²

¹Department of Industrial Management Engineering, Kyonggi University

²Korea Productivity Center

The diagnostic evaluation model for the smart factory operation system, established as a national standard (KS X 9001-3) in 2016, consists of 46 items in 10 categories. The level of smart factories has been quantitatively and qualitatively diagnosed based on KS for 49 companies in the manufacturing sector, and this paper analyzed the data of this diagnosis. It proposed the characteristics that could be based on the smart factory construction (specialization) project by industry, and the priority promotion and convergence improvement areas among the evaluation areas to improve the level of smart factories. This paper can be utilized in updating reference models by industry for the development of smart factories, improving the level diagnosis models of smart factories, establishing policies related to building smart factories, or establishing directions for improvement of companies.

Keywords: Smart Factory, LDA, Topic Modeling, Multiple Linear Regression

1. 서론

전통 제조업에 ICT를 결합하여 개별공장의 설비와 공정을 네트워크화하는 스마트공장은 모든 생산데이터가 실시간으로 공유 및 활용되어 최적화된 생산 운영이 가능한 공장으로서 공장 간 협업적인 운영을 가능하게 한다(Byun, 2016).

최근 제조업에서는 이러한 스마트공장 도입의 요구가 증가하고 있으며 스마트공장 구축의 필요성 및 관련 내용의 인지도가 높아지고 있다(Sung, 2019). 한국 정부 또한 제조업의 중요성을 인식하고 산업혁신 3.0의 개념을 도입하여 2019년 기준 총 12,660개의 스마트공장을 보급하였으며, 중소기업 대상 스마트공장 3만 개 보급(2022년) 및 스마트산단 20개 조성(2030년)을

목표로 중소·중견기업들에 스마트공장 보급을 위한 자금지원과 우수모델의 발굴 및 확산에 속도를 내고 있다.

국내 대기업들은 각종 데이터에 기초한 분석을 통해 실제 제조현장의 생산성 효율 및 품질개선을 오래전부터 실현해오고 있지만, 중소·중견기업들은 대기업의 1차 협력 업체를 제외하고는 많은 부분에서 제조현장의 스마트화가 부족하다(Lee *et al.*, 2018). 투자비용의 부담, 투자 대비 이익의 불확실성, 전문인력 부족 등과 같은 이유로 미도입 기업 혹은 낮은 수준의 스마트공장에 머물러있다(Chang *et al.*, 2019).

기업들은 개선을 위해 업종 및 수준별 요구사항이 기재된 ‘스마트공장 업종별 참조모델(KOIIA, 2017)’이나 KS X 9001-3 기반 스마트공장 수준 진단보고서의 구축전략 및 추진과제를

이 논문은 산업통상자원부의 재원으로 한국산업기술진흥원(KIAT)의 지원을 받아 수행된 연구임(2020년 스마트공장 운영설계 전문인력 양성사업, 과제번호 : N0002429).

* 연락저자 : 장태우, 16227 경기도 수원시 영통구 광고산로 154-42 경기대학교 산업경영공학과, Tel : 031-249-9754, Fax : 031-244-3534,

E-mail : keenbee@kgu.ac.kr

2020년 10월 19일 접수; 2020년 11월 26일 수정본 접수; 2020년 12월 31일 게재 확정.

참고할 수 있는데, 이들은 기업 활동의 전 영역을 아우르는 제안사항을 포함하고 있다. 그러나 투자비용의 부담을 느끼는 중소기업들은 모든 피드백을 수용하는 데 어려움이 있기 때문에, 핵심적 개선사항만을 우선 제시할 필요가 있다.

따라서 본 연구는 대기업과 비교하면 자체적 스마트공장 도입에 어려움이 있는 중소·중견기업을 대상으로 하여, 스마트수준 진단보고서에 명시된 진단데이터를 통해 최종 스마트수준 향상을 위한 업종별 스마트공장 목표 수준과 정책적 고려사항을 제시하고, 우선추진분야와 융합개선분야를 제시하고자 한다. 또한, 구축기업을 향한 스마트공장 건설턴트의 피드백을 통해 업종별 중점 토픽을 확인한 후, 업종별 개선기회를 제시하고자 한다.

본 연구의 분석 대상인 49개 기업에 대한 진단보고서에는 정량적인 데이터로 평가영역별 점수 및 스마트 수준 종합점수(이하 종합점수)와 정성적 데이터로 평가영역별 개선기회의 내용이 있다. 수준진단은 사전 인터뷰를 통한 기업의 요구사항 분석과 전문위원 2~3명으로 구성된 진단팀의 기업별 제조 현장진단으로 수행된다. 진단결과를 바탕으로 과제 실행 방안을 도출하고 컨설팅을 완료한다. 그러나 진단팀마다 수준 진단보고서를 작성하는 특성 때문에 기준이 상이한 개선기회가 도출될 수 있다.

진단보고서의 정량적 데이터로 다중회귀분석을 시행하여 스마트 수준에 가장 영향을 많이 미치는 평가영역을 도출하고 영역별 상관관계 분석을 시행해 융합적 개선이 가능한 영역을 도출하였다. 정성적 데이터인 진단보고서에 기술된 개선기회의 내용을 대표적 토픽 모델링 기법 LDA(Latent Dirichlet Allocation, 잠재 디리클레 할당)를 통해 업종별 토픽을 도출해 내고, 업종별 중점 토픽의 분포에 차이가 있음을 확인하였다. 이처럼 기업에 대한 스마트공장 진단보고서를 바탕으로 현 수준 진단모델 및 구축사업의 개선점과 방향성을 도출했다는 점에서 의의를 찾을 수 있다.

본 연구의 구성은 다음과 같다. 제 2장에서는 본 논문에서 사용한 데이터가 도출되는데 활용된 진단모델과 유의미한 결론 도출을 위한 연구방법론을 소개한다. 제 3장부터 제 5장까지는 업종별 스마트공장의 수준 분석과 개선기회 토픽모델링을 활용해 도출된 업종별 개선점, 기업 우선추진분야 및 융합개선분야 결과를 해석한다. 마지막으로 제 6장에서는 연구의 결론 및 시사/한계점을 언급한다.

2. 진단모델 및 연구방법론

2.1 스마트공장 진단평가모델

스마트공장 진단평가모델은 스마트공장의 수준 및 구성요소에 대해 민관합동스마트공장추진단과 한국생산성본부가 진행한 KS 규격화로 2016년 6월에 제정 고시되었다(Lee, 2018). 스마트공장 진단평가모델 개발의 목적은 (1) 제조업 스마트화 수준을 객관적으로 진단 및 평가하여 맞춤형 스마트공장 구축 방안 제시 및 인증을 위한 스마트 수준진단 도구 제공, (2) 규격화된 종합평가체계(시스템 구축 여부, 생산성 등)를 대기업 등 민간인증과 연계한 스마트공장 관련 표준모델 제시, (3) 스마트공장의 유효성을 객관적으로 입증함으로써 스마트공장 확산 기반을 조성하는 데 있다(Woo et al., 2018).

스마트공장 진단모델은 Goal & KPI(Key Performance Indicator) 달성을 위해 <Table 1>과 같이 4대 분야 10개 평가영역(모듈) 46개 세부 평가항목으로 설계되었다. 또한, 스마트공장 종합진단보고서는 진단모델에 기반을 두고 10개 평가영역에 대한 종합리포트 및 모듈별 진단결과, 개선 이슈를 도출하는 등 기업 제조역량에 대한 종합적이고 입체적인 진단결과를 제공한다.

Table 1. Smart Factory Evaluation Areas and Detailed Evaluation Items by Evaluation Area

Field	Evaluation Area	Detailed Evaluation Area
Strategy	1. Leadership & Strategy	Understanding Smart Factory, Strategy and Implementation Plan for Smart Factory, Organization and Capability Management for Smart Factory, Performance Indicators(KPI) Management
	2. Product Development	Product development procedure, Product design and verification, Factory Design and verification, Product information management, Technical information management
Process	3. Production Plan	Standard information management, Response to demand and orders, Mid-to Long-term, Production plans, Short-term production schedules
	4. Process Management	Establishment and Direction for work plans, Management for production progress, Response to abnormality
	5. Quality Control	Quality information management, Quality standards and Documents management, Inspection data management, Inspection equipment and Measurement equipment management
	6. Facility Management	Facility operation management, Facility conservation management, Maintenance and Repair management, Mold/Jig/Tools management
	7. Logistics Operation	Material requirement management, Order and Delivery date management, Incoming and Outgoing management, Loading and Storage, Pick-up and shipment management, Transportation/distribution information management
System & Automation	8. Information System	ERP(Enterprise Resource Planning)&SCM(Supply Chain Management) construction, MES(Manufacturing Execution System) construction, PLM(Product Life cycle Management) construction, FEMS(Factory Energy Management System) construction, Security management
	9. Facility Automation	Production facilities, Logistics facilities, Inspection facilities, Facility information network, Energy/Safety/Environmental management
Performance	10. Performance	Productivity, Quality, Cost, Delivery date, Safety, Environment

Table 2. Smart Factory Maturity Level and Characteristics

Level	Maturity Level	Characteristics
1	Checking	Simple state detected (not associated with external system)
2	Monitoring	The detection results are shown to the external monitoring system as data
3	Control	Information on detection results is analyzed with or without abnormalities and the function is normalized (compared to the reference information).
4	Optimization	Integrated information and comprehensive control from the perspective of system optimization through interfacing between internal and external systems between upper and lower facilities and equipment
5	Autonomy	Artificial intelligence and autonomous control of self-diagnosis and self-repair

<Table 2>는 제조시스템의 스마트 성숙도 수준을 평가하기 위한 척도이다. 성숙도 수준은 총 5수준으로 구성되었으며, 5점 만점을 기준으로 한다.

2.2 다중선형회귀분석 모델링

회귀분석은 반응(종속)변수에 대한 설명(독립)변수의 영향력을 예측하는 방법으로 반응변수 변화의 원인을 설명변수로 해석할 수 있다. 반응변수와 설명변수의 관계를 선형적이라고 가정하는 경우를 선형회귀분석이라 하며 이 모델은 표본데이터에 의한 과적합을 줄일 수 있다(Ryu *et al.*, 2016). 실제 데이터는 변수 하나만으로 회귀분석을 할 수 있는 경우가 매우 드물기 때문에 다음과 같이 다수(p)의 설명변수를 활용한 다중선형회귀분석이 주로 사용된다.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \epsilon$$

다중회귀분석 모델의 일반적인 형태는 위 수식과 같으며 잔차 ϵ 는 기댓값 0과 일정한 분산 σ^2 의 정규분포를 따른다고 가정한다. 선형회귀분석의 회귀계수를 추정하기 위해 최소제곱법을 사용하며 다음과 같은 오차제곱합이 최소가 되는 회귀계수를 추정한다.

$$L = \sum_{j=1}^n \epsilon_j^2 = \sum_{j=1}^n \{y_j - (\beta_0 + \beta_1 X_{1j} + \beta_2 X_{2j} + \dots + \beta_p X_{pj})\}^2$$

다중회귀분석에서 어떤 설명변수를 회귀식에 포함할 것인가는 중요한 문제 중 하나로서 설명변수를 많이 포함할수록 반응변수의 예측값의 분산과 자료수집 비용이 증가한다. 적절한 설명변수를 선택하는 방법은 대표적으로 3가지가 존재한다.

(1) 전진 선택법(Forward Selection)

설명변수가 아무것도 없는 모델부터 유의한 변수부터 하나씩 추가해 나가면서 더 이상 통계적으로 기여도가 유의하지 않다고 판단될 때 변수선택이 종료된다. 변수가 많을 때 빠르게 변수선택을 할 수 있는 장점이 있지만 선택된 변수는 다시 제거되지 않는다.

(2) 후진 제거법(Backward Elimination)

모든 설명변수가 포함된 모델에서 통계적으로 유의하지 않은 변수를 하나씩 제거해 나가면서 더 이상 제거할 회귀변수가 없다고 판단될 때 변수선택이 종료된다. 변수가 많을 경우 학습 초기에 계산시간이 길어 전진 선택법에 비해 느릴 수 있지만 필요한 변수를 제거의 대상에서 제외하며 변수를 선정하기에 유의미한 변수가 소거될 확률이 낮다.

(3) 단계적 선택법(Stepwise Selection)

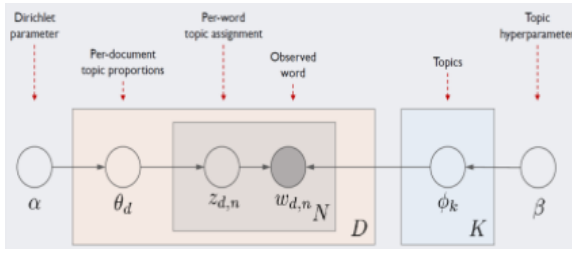
설명변수의 추가(Forward Selection)와 제거(Backward Elimination)를 병행하여 한 번 선택되거나 소거된 설명변수를 다시 고려하는 양방향 변수선택법이다. 한 번 선택되거나 소거된 변수가 고정되지 않기 때문에 결과가 비교적 안정적이거나 많은 경우의 수를 계산하는 데 더 긴 시간이 소요되어 학습이 느린 단점이 있다.

회귀분석은 설명변수가 서로 독립이라는 가정 하에서 이뤄지며 이때 각 설명변수의 설명력이 높다. 그러나 설명변수 간의 높은 상관관계가 존재할 경우 정확한 회귀계수 추정 및 검증의 변동성이 커지는 문제가 발생한다. 따라서 설명변수 간의 완전한 또는 거의 완전한 선형의 종속관계가 존재하는 것을 의미하는 다중공선성을 파악한 후 발생하지 않는 일부의 설명변수 집합을 추출해서 회귀모형을 구축해야 한다.

2.3 LDA 토픽 모델링

잠재 디리클레 할당(Latent Dirichlet Allocation, LDA)은 Blei, Ng, Jordan이 제시한 알고리즘으로 말뭉치의 생성 확률론적 모델이다. LDA는 문서에 대하여 어떤 주제들이 존재하는지를 확률로 표현한 모델로서, 각 주제는 단어 수의 분포를 기반으로 하는 특징을 갖는다(Blei *et al.*, 2003).

<Figure 1>과 같이 Dirichlet parameter α, β 를 이용하여 d 번째 문서의 초기 토픽 할당 분포인 θ_d 와 전체 말뭉치(corpus)의 초기 토픽 할당 분포인 ϕ_k 를 설정한다. d 번째 문서의 n 번째 단어에 할당된 토픽 $z_{d,n}$ 을 깃스샘플링(Gibbs Sampling) 방법을 통해 모든 단어에 대한 토픽 할당 정보가 수렴할 때까지 업데이트하여 ϕ_k 와 θ_d 의 사후 확률분포를 구한다(Noh, 2020).



D : Number of document
 K : Total number of topics
 N : The total number of words in the d th document
 θ : Topic Distribution per Document, $\theta_d \sim Dir(\alpha)$ for $d \in 1, 2, \dots, D$
 ϕ : Word Distribution per Topic, $\phi_k \sim Dir(\beta)$ for $k \in 1, 2, \dots, K$
 $z_{d,n}$: Topic distribution for that word, $z_{d,n} \sim Multi(\theta_d)$
 $w_{d,n}$: The n th word in the d th document, $w_{d,n} \sim Multi(\phi_{z_{d,n},n})$

Figure 1. Structure of LDA

LDA 토픽모델링을 위한 토픽 수 K 는 사용자가 지정하는 하이퍼파라미터로, 여러 번의 실험을 통해 도출해야 하는 값이며, 특정 확률모델이 실제로 관측되는 값을 얼마나 잘 예측하는지를 뜻하는 perplexity 값으로 도출한다. 이 값이 작을수록 토픽모델이 문서를 잘 반영한다고 할 수 있다(Eum et al., 2019).

$$Perplexity(w) = \exp \left[- \frac{\log \{p(w)\}}{\sum_{d=1}^D \sum_{j=1}^V n^{jd}} \right]$$

3. 업종별 스마트공장 수준 분석

3.1 데이터 분류

업종 구분을 위해 수준 진단된 49개 기업의 업종을 한국표준산업분류체계 10차 개정안을 참고하여 각각 중분류, 소분류, 세분류까지 할당하였다. 중분류를 기준으로 12개 업종, 소분류를 기준으로 19개 업종으로 구분되었다. 업종의 대분류가 제조업 하나로 통일되어 있고 중분류별 기업의 수가 많지 않음을 감안하여, 분석의 구체화를 위해 <Table 3>과 같이 임의로 금속, 기계, 섬유화학, 전자, 선박, 자동차의 6개 업종으로 구분하였다.

3.2 업종별 평가지표 점수 시각화

전반적인 데이터 수준 분포를 확인하기 위해 업종 중분류별 평가영역 평균 수준 점수를 활용하여 <Figure 2>와 같이 히트맵으로 표현하였으며, 고무 및 플라스틱 제품 제조업(C22), 1차 금속 제조업(C24), 전자 부품, 컴퓨터, 영상, 음향 및 통신장비 제조업(C26)이 전반적으로 우수한 점수를 받은 업종으로 나타났다. 또한, 업종과 상관없이 경영 관점의 ‘1. 리더십&전략, 10. 성과’ 영역보다 관리 및 운영 관점 평가영역의 점수가 낮았으며 특히 ‘6. 설비관리’ 영역의 평균 수준이 1.18로 가장 낮게 나타났다.

구체적으로 드러나는 특징을 도출하기 위해 ‘업종별로 평가영역 수준에 차이가 있다’라는 가정하에 임의로 구분한 6개 업종별 데이터와 전체 데이터의 평가영역별 평균 수준 차이를 <Figure 3>과 같이 막대그래프로 나타내었다. 선박업종은 전 영역에서 편차가 양수로 평가영역별 평균점수가 가장 우수하였고, 반대로 기계업종의 평가영역별 평균점수가 가장 낮게 나타났다. 섬유화학업종 또한 다 영역에서 미비한 수준으로 나타났다.

<Figure 4>는 스마트수준이 가장 낮은 기계업종의 전체평균과 평가지표별 평균 차이를 나타낸 것으로, 기계업종의 전체 평균 수준(1.5)과 비교하여 평가지표별 편차를 확인한 결과, 1. 리더십&전략, 8. 정보시스템, 10. 성과 영역을 제외한 항목에서 평균 수준보다 낮음을 확인하였다. 특히 <Figure 3>과 비교하여 타 업종보다 5. 품질관리와 6. 설비관리 영역에서의 수준이 낮음을 확인하였다.

Table 3. Distribution of Companies by Industry

Industry	Code	Mid Category	Company
Metal	C24	Manufacture of basic metals	3
	C25	Manufacture of fabricated metal products, except machinery and furniture	4
Machine	C29	Manufacture of other machinery and equipment	13
Fiber	C10	Manufacture of food products	1
	C13	Manufacture of textiles, except apparel	4
Chemistry	C20	Manufacture of chemicals and chemical products; except pharmaceuticals and medicinal chemicals	5
	C22	Manufacture of plastics and synthetic rubber in primary forms	2
Electronics	C26	Manufacture of electronic components, computer; visual, sounding and communication equipment	2
	C27	Manufacture of medical, precision and optical instruments, watches and clocks	2
	C28	Manufacture of electrical equipment	3
Automobile	C30	Manufacture of motor vehicles, trailers and semitrailers	7
Ship	C31	Manufacture of other transport equipment	3

Code	Industry	1	2	3	4	5	6	7	8	9	10
C10	Manufacture of food products	2.63	1.20	1.63	1.00	1.38	1.00	1.33	1.40	1.10	2.00
C13	Manufacture of textiles, except apparel	1.88	1.38	1.69	1.67	1.41	1.13	1.25	1.40	1.40	2.13
C20	Manufacture of chemicals and chemical products; except pharmaceuticals and medicinal chemicals	2.09	1.42	1.46	1.03	1.25	1.05	1.36	1.42	1.29	2.30
C22	Manufacture of plastics and synthetic rubber in primary forms	2.50	1.40	1.56	1.83	1.50	1.47	1.63	2.15	1.40	2.33
C24	Manufacture of basic metals	2.08	1.60	1.48	1.89	1.69	1.63	1.64	1.65	1.80	2.67
C25	Manufacture of fabricated metal products, except machinery and furniture	1.94	1.70	1.53	1.33	1.41	1.19	1.41	1.58	1.28	2.25
C26	Manufacture of electronic components, computer, visual, sounding and communication equipment	2.38	2.60	2.38	1.33	1.75	1.38	2.00	2.10	1.40	2.00
C27	Manufacture of medical, precision and optical instruments, watches and clocks	2.19	1.83	1.94	1.21	1.56	1.25	1.46	1.65	1.53	2.31
C28	Manufacture of electrical equipment	2.17	1.30	1.17	1.11	1.33	1.00	1.31	1.50	1.30	2.31
C29	Manufacture of other machinery and equipment	2.20	1.42	1.47	1.27	1.13	1.04	1.41	1.52	1.25	2.31
C30	Manufacture of motor vehicles, trailers and semitrailers	2.10	1.63	1.55	1.27	1.54	1.34	1.37	1.64	1.43	2.36
C31	Manufacture of other transport equipment	2.23	1.68	1.54	1.36	1.40	1.31	1.58	1.77	1.40	2.36
Total Average		2.14	1.53	1.54	1.34	1.36	1.18	1.42	1.58	1.36	2.31

Figure 2. Heat Map of Diagnosis Scores by Industry and Evaluation Area

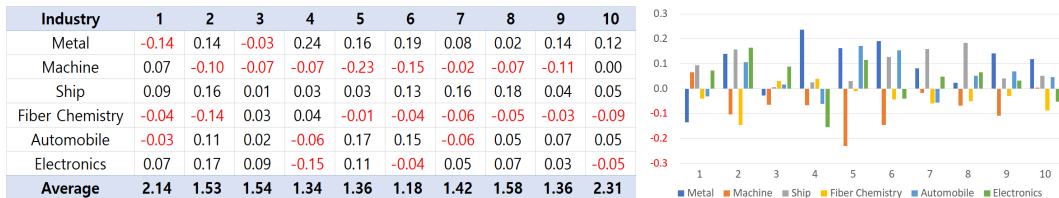


Figure 3. Mean Deviation of Evaluation Area Score by Industry

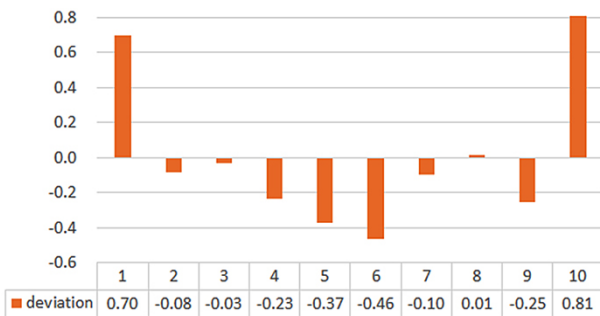


Figure 4. Difference between Industry Average and Evaluation Index Average

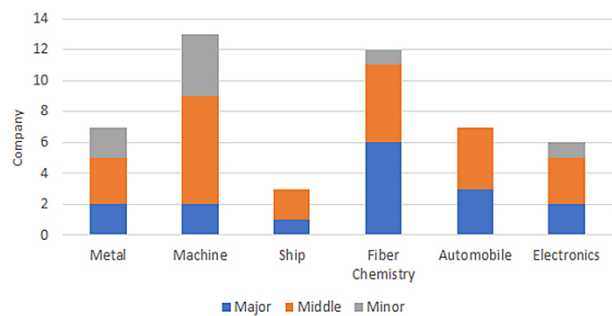
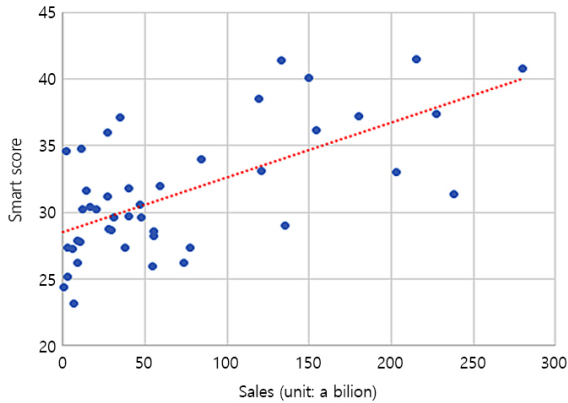


Figure 5. Number of Companies by Industry Type and Company Size

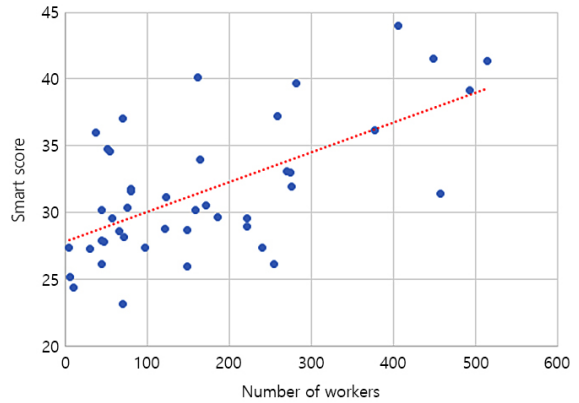
3.3 기업 규모의 영향성 파악

3.2절에서 나타난 특징들이 기업규모의 영향을 받아 나타났을 가능성을 파악하기 위해 <Figure 5>와 같이 업종별 기업규모의 비율을 도출하였다. 기업의 규모는 중소기업벤처부에서 발행한 중소기업 범위해설집 내 업종별 규모기준을 참조하였으며, 매출액을 기준으로 대, 중, 소로 분류하였다. 평균 종합점수가 낮은 기계, 섬유화학 중 기계업종에서는 규모가 중, 소인 기업 수가 많음을 확인하였다.

구체적인 규모의 영향성 파악을 위해 기업규모와 종합점수의 상관관계를 매출액과 종합점수 관계, 근로자 수와 종합점수의 관계 두 경우로 관계를 파악하였다. 아웃라이어는 각각 5개 (매출액 3천억 원 이상, 근로자 수 600명 이상 기업)를 제외하고 상관분석을 실시하였다. 매출액과 종합점수, 근로자 수와 종합점수 두 경우의 상관계수가 각각 0.65, 0.62로 기업규모와 스마트수준이 대체로 비례하는 것으로 나타났다. 각각 44개 진단기업의 분포는 <Figure 6 (a)>, <Figure 6 (b)>과 같다.



(a) Sales and Smart score



(b) Number of workers and Smart score

Figure 6. Correlation with Smart Score

3.4 업종별 분석 종합 및 제언

업종별, 규모별, 평가영역별을 엑셀을 활용하여 간단히 분석해본 결과 분석에 활용된 6가지 업종 중 기계업종의 수준이 가장 낮음을 확인할 수 있었으며, 대부분의 평가지표영역 중 1. 리더십&전략 및 10. 성과 영역을 제외한 8가지 영역의 평균 수준이 1.5 내외인 것으로 나타났다. 업종별로 분류된 기업규모를 파악해보았을 때, 중소기업의 수가 다른 업종에 비해 많았던 기계업종의 수준이 낮았으며, 상관분석을 통해 스마트 수준이 규모에 영향을 많이 받음을 확인하였다. 따라서 스마트공장 지원사업을 시행하는 기업은 위와 같은 분석결과를 참고하여 업종뿐만 아니라 규모에 따른 지원사업, 특히 중소기업 맞춤형 지원사업과 같은 특화사업에 중점을 둘 필요가 있다.

또한, 임의로 구분한 업종 내에서도 동일업종이지만 회사의 특성(뿌리기업 혹은 가공기업)에 따라서 프로세스의 차이가 존재한다. 따라서 그에 맞춰 지원사업이 더 세부적으로 구성될 필요가 있다.

4. 토픽모델링을 활용한 업종별 개선점 도출

4.1 업종별 토픽 도출

스마트공장 수준 진단보고서에서는 수준 진단뿐만 아니라 평가영역별로 개선기회가 제시되어 있다. 그러나 이 개선기회를 도출하는 과정에서 컨설팅을 추진한 기업 또는 2~3명의 컨설턴트로 구성된 진단팀마다 기준이 다른 개선 방향을 제시할 수 있다. 이에 스마트공장 수준 개선을 위해 컨설턴트들이 제시한 전반적인 개선방향을 파악하고자 하였다. 분석은 평가영역이 다른 1개의 보고서를 제외한 48개 보고서를 대상으로 하였고, 토픽모델링 분석을 통해 업종별 중점영역과 차이점을 살펴보았다.

<Table 4>는 업종별 보고서 수 및 제시된 개선기회의 수*를 나타낸 것으로, 수준이 낮았던 기계업종과 섬유화학업종의 개선기회 수가 타 업종에 비해 많은 것을 알 수 있다.

Table 4. Number of Reports by Industry and Number of Opportunities for Improvement

Industry	Number of Reports	Opportunity for Improvement
Metal	7	442
Machine	13	1132
Ship	3	158
Fiber Chemistry	12	1117
Automobile	7	452
Electronics	6	692
Total	48	4003

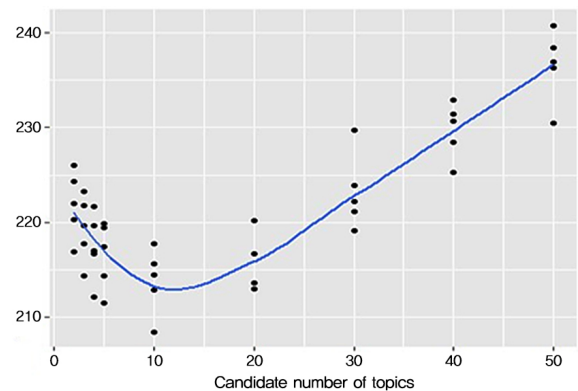


Figure 7. Perplexity for Determining the Number of Topics

R 프로그램을 이용하여 LDA 토픽모델링 분석을 실시하였으며, 사용된 패키지는 KoNLP, tm, topicmodels, lda, slam, doppel, parallel, ggplot2, LDAvis, stringr, servr, MASS이다.

적절한 토픽의 수를 결정하기 위해 <Figure 7>과 같이 모든 데이터가 최소 한번은 test set으로 쓰이도록 하는 5-fold Cross Validation을 활용하여 perplexity 값을 도출하였다. <Figure 7>에서 x축은 토픽의 수, y축은 perplexity 값을 의미한다. 토픽 수 K는 perplexity의 값이 가장 작은 12로 지정하였으며, 토픽별로 10개의 단어를 뽑아 토픽을 구성하도록 설정하였다.

* 개선기회 수는 보고서별 제시된 개선기회의 업종별 합계.

<Table 5>와 같은 결과가 도출되었으며, 업종별로 어느 토픽에 중점을 두고 있는지 파악하기 위해, 해당하는 토픽 내 단어의 출현빈도를 업종별로 <Table 6>과 같이 나타내었다. 금속업종에서는 토픽 V7에 해당하는 단어가 가장 많이 출현

하였으며, 기계업종에서는 V11, 선박업종에서는 V6, 섬유화학업종에서는 V11, 자동차업종에서는 V12, 전자업종에서는 V11에 해당하는 단어가 가장 많이 나왔다. 예를 들어, 토픽 V1에 해당하는 단어가 1개라도 있는 금속업종의 개선기획의 수는 35개이다.

Table 5. Topics Derived from Main Terms

Topic	Main Terms	Derived Topics
V1	management, process, ERP, utilization, activity, unit, system, mold, segment, enhancements	ERP-centric process management activity
V2	KPI, inspection, preservation, competency, reflection, technology indicators, input, mid- to long-term, program	KPI-oriented inspection and conservation management
V3	record, standard, automation, strategy, condition operation, criteria, order, research, SCM	Record and Standard management strategy and SCM
V4	system, propulsion, interlocking, BOM, organization aggregation, inspection, layout, mobile, inside	BOM interlocking System
V5	process, system, barcode, necessary, operation, production plan evaluation, warehousing, manager, model	Process Operation System Using Barcode
V6	monitoring, occurrence, integration, development, foresight establishment, execution, key, training, status	Predictive preservation monitoring and training
V7	real-time, data, linkage, progress, line, introduction acquisition, customer, level, addition	Link between real-time data
V8	monitoring, quality, information, analysis, planning scheduling, changing, vision, packaging, installation	Vision-driven quality monitoring
V9	facility, MES, product, time, anomaly, phase, base, purchase, partner, APS	MES-centric facility abnormality management
V10	system, inventory, performance, field, automatic, utilization management, material, registration, PLC	Material Inventory Management and System Control
V11	automatic, construction, task, system, registration, design, project, improvement, organization, collection	Automated data collection and Construction of registration system
V12	function, material, production, work, defect instruction, order, element, goal, failure	Strengthen production capabilities to achieve objectives

Table 6. Number of Documents by Industry

Topic/Industry	Metal	Machine	Ship	Fiber Chemistry	Automobile	Electronics
V1	35	96	16	317	37	54
V2	22	94	16	335	49	58
V3	28	67	15	283	31	55
V4	36	89	10	325	46	58
V5	50	92	5	318	41	53
V6	27	109	22	358	33	77
V7	51	87	14	342	34	50
V8	39	80	13	326	35	55
V9	41	86	11	299	25	38
V10	38	98	12	323	32	48
V11	38	134	15	408	35	83
V12	39	102	11	369	56	63

4.2 업종별 개선기회 종합 및 제언

컨설턴트가 제시한 개선기회는 해당 기업이 나아가야 할 개선점임과 동시에 컨설턴트들의 평가 기준이 반영된 것으로 볼 수 있다. 비록 서로 다른 기업 및 컨설턴트의 개선기회라 할지라도, 기업의 스마트 수준 종합점수에 큰 차이가 없다는 가정 하에 비슷한 개선기회가 제시될 수 있다. 따라서 진단보고서에 기술된 개선사항 내용을 바탕으로 LDA 토픽모델링을 실시하여 나온 결과 또한 48개 기업에 대한 전반적인 평가라 볼 수 있다.

업종별로 어느 토픽에 중점을 두고 있는지 파악한 결과, 종합점수가 낮았던 기계, 섬유화학, 전자업종이 데이터 자동수집 및 등록시스템 구축과 관련한 토픽(V11)이 중요함을 알 수 있었다. 더불어 한 제품에 필요한 비용이 가장 큰 선박업종이 예지보전 모니터링 및 교육(V6)에 중점을 둔다는 점, 제품 제작에 부품이 다량 필요한 자동차업종은 ATO(Assemble to order) 생산이 주를 이뤄 납기준수와 같은 목표달성을 위한 생산업무 기능 강화(V12)에 중점을 둔다는 점, 온·습도 및 자재 함량에 민감한 금속업종은 실시간 데이터 연계(V7)에 중점을 둘 필요가 있다는 점을 제시하고 있다고 추정할 수 있다. 스마트공장 지원 사업을 시행하는 기업들은 이러한 결과를 참고하여 업종별 지원 내용에 중점을 둘 필요가 있다.

그러나, 종합진단 보고서의 개수가 적고, 영역별 개선기회의 양이 적은 것이 한계라 볼 수 있다. 종합진단 보고서가 축적된다면, 보다 의미 있는 결과가 도출될 것이다.

5. 기업 우선추진분야 및 융합개선분야 분석

5.1 기업별 우선추진분야 제시

스마트공장 영역별 전략을 제시하여 기업들이 진단모델을 효과적으로 활용하도록, R프로그램을 이용하였으며, psych, car 패키지를 활용하여 단변량 다중회귀분석 모델 기반 변수 중요도를 탐색하였다. 단변량 다중회귀분석 모델은 두 개 이상의 독립변수(X)와 종속변수(Y)간 영향 관계를 설명하고, 도출된 회귀모델 식을 이용하여 종속변수 값을 예측하는 데 활용된다. 따라서 다중회귀분석을 통해 스마트 종합점수(Y)에 중점적으로 영향을 미치는 평가영역점수(X)를 도출한 후, 기업들이 우선으로 고려할 발전 영역을 제시하고자 한다.

(1) 분석 절차

<Table 7>과 같이 독립변수 X는 10개 부문 평가영역 점수, 종속변수 Y는 100점 기준 기업의 종합점수로 지정하였으며 종속변수는 연속형 범주이다. 49개 기업 중 스마트 종합점수가 누락된 기업 및 평가영역 점수에 N/A 값이 있는 9개 기업의

Table 7. Variable Definition

Variable	Variable Description
Y	Smart Factory Level Overall Score
X1	Leadership & Strategy Score
X2	Product Development Score
X3	Production Plan Score
X4	Process Management Score
X5	Quality Control Score
X6	Facility Management Score
X7	Logistics Operation Score
X8	Information System Score
X9	Facility Automation Score
X10	Performance Score

데이터를 제외한 총 40개 기업을 대상으로 분석을 진행하였다. 유의변수 선정에는 설명변수가 아무것도 없는 모델부터 유의한 변수부터 하나씩 추가해 나가는 전진선택법을 활용하였다. ANOVA로 전체 설명변수가 포함된 모델과의 설명력 차이를 비교하고 독립변수들 간에 상관관계를 확인하여 다중공선성이 존재하지 않는 최종 회귀모델을 구축하였다.

(2) 분석 결과

전진선택법을 이용한 회귀모델 구축 및 분석 결과, 종합점수에 주로 영향을 미치는 유의한 평가영역은 10개의 평가영역 중 3. 생산계획, 6. 설비관리, 9. 설비자동화, 10. 성과 영역으로, p-value가 0.05보다 작은 것으로 확인되었다. 또한, 회귀모델의 R² 값이 0.84로 84%의 설명력을 지니고 있으며 전체 변수가 포함된 모델과 비교해 보았을 때 설명력의 차이가 존재한다. 최종 회귀모델은 독립변수 간에 강한 상관관계가 나타나서 유의한 변수도 유의하지 않다고 결론을 도출할 확률이 생기기 때문에 다중공선성이 존재하지 않아야 한다. 유의한 변수들의 분산팽창요인(VIF, Variance Inflation Factors)값이 10을 넘으면 다중공선성에 문제가 있는 것으로 판단하지만, 본 회귀모델은 <Table 8>과 같이 VIF 값이 10보다 작으므로 다중공선성이 존재하지 않는 것을 알 수 있다.

따라서 주요요인을 ‘3. 생산계획, 6. 설비관리, 9. 설비자동화, 10. 성과’로 도출할 수 있었다. 스마트제조혁신추진단에서 발표한 스마트 수준 자가진단지에서도 평가영역별로 가중치가 다른 것을 감안할 때, 높은 스마트 수준을 얻으려는 목표를 가진 스마트공장 미구축 기업이나 도입을 했음에도 불구하고 기초수준(1~2수준)인 기업들에 네 가지 평가영역(3. 생산계획, 6. 설비관리, 9. 설비자동화, 10. 성과)의 우선 관리체계를 구축한다면 향후 기업들의 빠른 스마트공장 수준 향상을 도모할 수 있을 것이다.

Table 8. Detecting Multicollinearity Using Variance Inflation Factor

	3. Production Plan	6. Facility Management	9. Facility Automation	10. Performance
VIF	1.18	1.60	1.58	1.07

예를 들어, A기업의 목적이 스마트수준 향상이라면 4개 영역의 일부인 6. 설비관리 영역의 개선을 강조할 수 있으며, 강조 방식으로는 세부 평가항목의 설비가동관리를 평가할 수 있는 설비성능가동률을 핵심성과지표로 설정한 후 표준 사이클타임을 집중적으로 관리하도록 지시할 수 있다.

설비 성능 가동률

$$= \frac{\text{생산수량} \times 1\text{개 생산량의 표준사이클타임}}{\text{가동시간}} \times 100\%$$

5.2 융합개선분야 제시

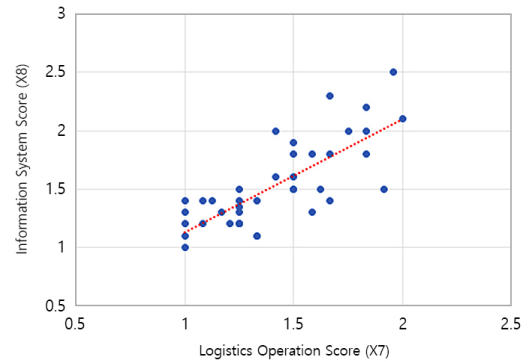
관리영역이 연관됨으로써 융합개선의 기회가 있는 분야를 찾아 기업들에게 제시할 필요가 있다. 이를 위해 R 프로그램을 이용하여 변수 상관관계 분석을 시행하였다. <Table 9>와 같이 평가영역 중 물류운영(X7)과 정보시스템(X8), 제품설계(X2)와 정보시스템(X8), 제품설계(X2)와 물류운영(X7)이 각각 0.77, 0.76, 0.73으로 강한 상관관계가 나타났다. 각 기업의 두 점수를 시각화하여 나타내면 <Figure 8(a)>, <Figure 8(b)>, <Figure 8(c)>와 같다.

5.3 개선영역 분석 종합 및 제언

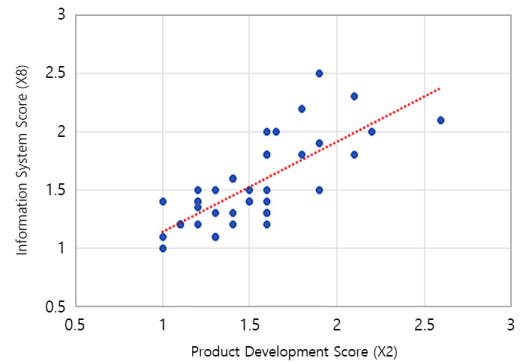
스마트공장 시스템 도입의 가장 큰 우려사항은 투자비용에 대한 부담이며(Chang *et al.*, 2019), 중소기업들은 특히나 스마트공장 구축사업 진행 시 투자의 동반이 필수적이다. 본 분석은 주요 요인분석을 시행하여 스마트 수준에 가장 영향을 많이 미치는 평가영역을 도출하고 영역별 상관관계 분석을 시행해 융합적 개선이 가능한 영역을 도출하였다. 분석을 통해 도출된 영역을 바탕으로 우선적인 개선을 시행한다면, 적은 노력으로 스마트 수준을 빠르게 향상할 수 있다. 향후 이러한 분석결과를 참고하여 영역별 투자비의 차이를 두거나 집중 투자와 같은 전략을 가진 스마트공장 구축사업을 진행한다면 투자 대비 효과의 명확한 확보를 보장하는 사업이 기대된다.

Table 9. Correlation between 10 Evaluation Area Scores

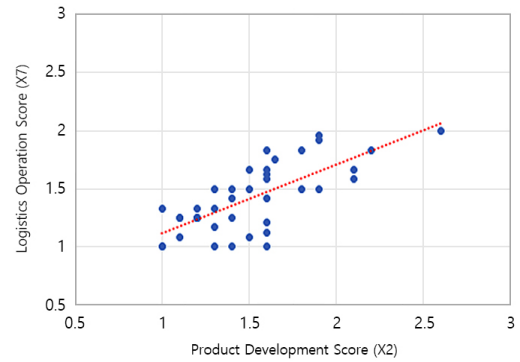
	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
X1	1.00									
X2	0.29	1.00								
X3	0.30	0.65	1.00							
X4	0.14	0.28	0.35	1.00						
X5	0.14	0.67	0.57	0.37	1.00					
X6	0.06	0.53	0.38	0.47	0.51	1.00				
X7	0.24	0.73	0.55	0.46	0.47	0.58	1.00			
X8	0.47	0.76	0.69	0.53	0.56	0.55	0.77	1.00		
X9	0.22	0.46	0.30	0.37	0.42	0.57	0.50	0.53	1.00	
X10	0.34	0.20	0.05	0.06	0.02	0.15	0.25	0.22	0.26	1.00



(a) X7 and X8



(b) X2 and X8



(c) X2 and X7

Figure 8. Strong Correlation between Evaluation Areas

6. 결 론

본 논문은 KS X 9001-3의 진단평가모델을 기반으로 한국생산성본부에서 진행된 49개 기업에 대한 진단보고서를 활용하여 연구를 진행하였고, 총 3가지 분석을 진행하였다.

첫째, 기업별 스마트화 수준 데이터를 업종 및 규모별로 구분하여 분석을 진행하였다. 스마트공장 구축현황의 업종별 차이, 평가영역 수준의 차이, 스마트 수준과 규모의 연관성을 확인하였다. 기업별 평가영역 점수를 비교한 결과 1. 리더십&전략 영역과 10. 성과 영역이 가장 높은 수준임을 파악하였고, 나머지 8가지 영역 모두 1.5 수준 내외로 낮은 점수를 보인 것을 알 수 있었다. 또한, 타 업종에 비해 중소기업 수가 많았던 기계업종의 수준이 대체로 낮았으며, 스마트화 수준이 기업 규모에 영향을 많이 받았음을 상관분석을 통해 확인하였다. 둘째, 기업별 진단보고서에 기술된 개선기회를 데이터로 토픽모델링을 진행하여 12개의 주요 토픽을 도출하였으며, 업종별 중점 토픽의 분포에 차이가 있음을 확인하였다. 이처럼 상관관계분석과 토픽모델링을 통해 도출된 결과는 스마트공장 지원사업을 시행할 때 지원받는 기업의 규모 및 업종을 고려한 사업을 진행해야 함을 시사한다.

셋째, 앞서 활용한 기업별 평가지표를 다중회귀분석을 이용하여 통계분석을 시행하였다. 스마트 수준에 가장 영향을 미치는 평가영역을 확인하고 융합적 개선이 가능한 영역을 도출하고자 하였으며, 분석결과, 네 가지 영역(3. 생산계획, 6. 설비관리, 9. 설비자동화, 10. 성과)이 주로 스마트 수준에 영향을 미침을 확인하였다. 또한, 상관관계분석을 통해 7. 물류운영과 8. 정보시스템, 2. 제품설계와 8. 정보시스템, 2. 제품설계와 7. 물류운영이 상관관계가 있음을 알 수 있었다. 이러한 결과를 통해 도출된 우선추진분야 및 융합개선분야는 적은 노력으로 스마트 수준을 빠르게 향상할 수 있음을 시사한다.

본 논문은 기업에 대한 스마트공장 진단보고서를 바탕으로 위와 같은 현 수준 진단모델 및 구축사업의 개선점과 방향성을 도출했다는 점에서 의의를 찾을 수 있으며 스마트공장 발전을 위한 업종별 참조모델 갱신, 스마트공장 수준 진단모델의 개발, 스마트공장 구축 관련 정책 또는 기업의 개선 방향 수립 등에 활용될 수 있을 것이다. 그러나 연구결과를 도출하는데 사용된 진단보고서의 수가 적기에 제조업 전반을 아우르는 결과라고 확인할 수 없다. 향후 많은 보고서 내용이 축적되면 더 의미 있는 결과의 도출이 가능할 것이라 기대된다. 또한, 현재 스마트공장 진단평가모델은 평가체계가 컨설턴트를 통한 인터뷰 및 현장투어로 진행되기 때문에 평가 자체가 주관적이라는 한계점을 가지고 있어 수준확인제 등을 고도화할 필요가 있다.

참고문헌

Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003), Latent Dirichlet Allocation, *Journal of Machine Learning Research*, **3**, 993-1022.
Byun, D. H. (2016), Trend of Smart Factory and Model Factory Cases, *The e-Business Model*, **17**(4), 211-228.

Chang, T.-W., Sung, S.-I., and Lee, J.-C. (2019), Survey Analysis on Small and Medium-Sized of Smart Factory and Improvement Plan for Them, *Entrue Journal of Information Technology*, **17**(1), 77-88.
Eum, S.-M., Lee, S.-G., Cho, S.-W., and Lee, C.-W. (2019), Analysis of Research Trends of Wireless Power Transfer System for Locomotives Using Topic Modeling Based on LDA Algorithm, *Journal of Korean Institute of Industrial Engineers*, **45**(4), 284-301.
KOIIA (2017), Smart Factory Reference Model Industry-Oriented.
Lee, H.-Y. and Chang, J.-Y. (2018), Analysis and Response Strategies of Small and Medium-Sized Smart Factory in South-East Region, *The Journal of The Korean Institute of Communication Sciences*, **35**(4), 35-42.
Lee, J. C. (2018), Smart Factory Introduction Stage Propulsion Point, *Industrial Engineering Magazine*, **25**(4), 31-37.
Noh, S.-H. (2020), Analysis of Issues Related to Artificial Intelligence Based on Topic Modeling, *Journal of Digital Convergence*, **18**(5), 75-87.
Ryu, N.-H., Kim, H.-S., and Kang, P.-S. (2016), Evaluating Variable Selection Techniques for Multivariate Linear Regression, *Journal of the Korean Institute of Industrial Engineers*, **42**(5), 314-326.
Sung, C. Y. (2019), A Study on the Actual Condition of Construction in Smart Factory by Small and Medium-sized Manufacturing Companies, *Journal of the Korea Academia-Industrial Cooperation Society*, **20**(9), 182-187.
Woo, J.-H., Jeong, H., and Kim, Y.-M. (2018), A Study on the Diagnosis of Smart Factory for the Construction of Smart Shipyard, *Bulletin of the Society of Naval Architects of Koera*, **55**(4), 16-23.

저자소개

이민주 : 경기대학교 창의공과대학 산업경영공학과 학사과정 학생으로 재학 중이다. 관심분야는 스마트공장, 데이터마이닝이다.

이지수 : 경기대학교 창의공과대학 산업경영공학과 학사과정 학생으로 재학 중이다. 관심분야는 스마트공장, 데이터마이닝이다.

최지원 : 경기대학교 창의공과대학 산업경영공학과 학사과정 학생으로 재학 중이다. 관심분야는 스마트공장, 데이터마이닝이다.

장태우 : 서울대학교 산업공학과에서 학사를 마쳤으며, 동 대학원에서 산업공학 석사, 박사학위를 취득하였다. 한국전자통신연구원에서 우편 정보화/자동화를 연구하였고, 현재 경기대학교 산업경영공학과에 재직 중이다. 주요 연구 관심분야는 스마트공장, 물류/SCM, 시스템분석 등이 있으며, 현재 경기도 지역협력연구센터(GRRC)인 지능정보융합제조연구센터의 센터장을 맡고 있다.

배재현 : 국민대학교에서 전자공학 학사, 전과공학 석사학위를 취득하였고, 스마트공장 및 인공지능 분야의 컨설팅 & 교육분야 전문가로 재직 중이다. 연구분야는 스마트제조, 제조DX, 인공지능이다.