

데이터 마이닝 기법을 활용한 스마트팩토리 도입 기업의 특성 분석

오정윤¹, 최상현^{2*}

¹충북대학교 경영정보학과, ²충북대학교 경영정보학과

An Analysis of the Characteristics of Companies introducing Smart Factory System Using Data Mining Technique

Jeong-yoon Oh¹, Sang-hyun Choi^{2*}

¹Dept. Management Information Systems, Chungbuk National University

²Dept. Management Information Systems, Chungbuk National University

요 약 현재 스마트팩토리에 관한 연구는 구축 방안이나 설립 시 고려사항 등에 대해 꾸준히 진행되고 있다. 그러나 스마트팩토리를 도입한 기업에 대해서는 다양한 연구가 이루어지지 않고 있다. 이 연구에서는 스마트팩토리의 기초단계를 적용한 중소기업을 대상으로 설문조사를 실시하였다. 만족도의 특성을 확인하기 위해 군집분석을 하였고, 만족도에 따라 어떠한 특성을 가지는지 확인하기 위해 의사결정나무와 나이브베이즈 분석을 하였다. 군집분석 결과 만족도가 높은 그룹과 낮은 그룹으로 나뉘는 것을 확인하였으며, 의사결정나무와 나이브베이즈 분석을 실시한 결과 만족도가 높을수록 생산성 개선 정도가 높은 것을 확인하였다.

주제어 : 스마트팩토리, 군집분석, 의사결정나무, 나이브베이즈, 만족도

Abstract Currently, research on smart factories is steadily being carried out in terms of implementation strategies and considerations in construction. Various studies have not been conducted on companies that introduced smart factories. This study conducted a questionnaire survey for SMEs applying the basic stage of smart factory. And the cluster analysis was conducted to examine the characteristics of the company. In addition, we conducted Decision Tree and Naive Bay to examine how the characteristics of a company are derived and compare the results. As a result of the cluster analysis, it was confirmed that the group was divided into the high satisfaction group and the low satisfaction group. The decision tree and the Naive Bay analysis showed that the higher satisfaction group has high productivity.

Key Words : Smart Factory, Cluster, Decision Tree, Naive Bayes, Questionnaire

1. 서론

우리나라의 근대화를 이끌었던 제조업은 효율성 저하, 인구의 감소 및 노령화, 무역규모 감소, 공장의 해외 이전 등으로 위기를 겪고 있다. 주력 상품들이 성숙기에 접어들면서 후발 국가들과의 차이가 좁혀졌고, 제조업 발전

으로 안정기에 접어든 중국의 압박을 받게 되면서 ‘제조업 위기설’이 나오고 있다. Deloitte Global과 미국경쟁력위원회(US Council on Competitiveness)가 공동으로 발표한 ‘글로벌 제조업 경쟁력 지수 보고서’에서 2020년에는 한국이 6위로 한 계단 하락하고 미국은 1위를 탈환하며 중국은 1위에서 2위로 하락한다고 발표하였다. 또한

*We appreciate Sangsan Brick Co. supporting this research by the development fund of the Chungbuk National University.

*Corresponding Author: Sang-Hyun Choi(chois@cbnu.ac.kr)

Received March 6, 2018

Accepted May 20, 2018

Revised April 3, 2018

Published May 28, 2018

동남아 개도국들은 전반적으로 순위가 상승할 것이라고 발표하였다. 이는 한국이 제조업 정점을 찍고 하락한다는 의미이며, 후발주자인 중국과 동남아 개도국은 급부상하고 있다는 의미이다[1].

주요 선진국들은 제조업의 위기에 대응하고 중요성에 주목하기 위해 새로운 혁신 정책을 발표하였다. 독일은 2012년에 'Industry 4.0'을 발표하며 국가 10대 미래전략의 일환으로 민·관·학 연계를 통한 제조업 혁신을 추진하고 제조업과 ICT 융합을 통한 스마트 공장을 구축한다고 하였으며 EU는 2013년에 'Factories of the Future'를 발표하여 2020년까지 역대 제조업 비중을 향상시키고 사물인터넷, 가상현실 등에 기반해 전 제조공정의 유연화 및 네트워크를 실시한다고 하였다. 미국은 2009년 'Remaking America'를 발표하며 3D프린팅 등 첨단 제조기술 혁신과 산업용 로봇 활성화를 추진한다고 하였으며 일본은 2013년 '산업재흥플랜'을 발표하며 자동운전시스템 등의 차세대 인프라 구축에 힘쓴다고 하였다[2].

우리나라도 산업통상자원부에서 '제조업 혁신 3.0'을 발표했다. 계속되는 제조업 위기론에 대처하기 위해 정부는 '스마트 공장 추진단'을 구성하여 IT·SW·사물인터넷 등 기술 수준이 부족한 중소·중견 기업들을 대상으로 공장의 스마트화를 지원하고 있다. 민관협동 스마트공장 추진단은 급변하는 수요에 신속하게 대응하기 위하여 스마트공장 구축을 통한 맞춤형 유연생산 체제로의 전환은 필수적이라 밝혔다[3]. 2016년 9월 말 기준으로 2,611개사의 스마트공장 구축을 지원하여 불량률 감소, 원가 절감, 유연생산을 통한 생산품목 다양화 등 제조업 경쟁력 향상을 위한 의미있는 성과를 창출했으며, 2017년 중에 민관합동으로 총 1,108억원을 투입하여 2,200개 이상 중소·중견기업의 스마트공장 추진을 지원한다고 계획하였다[5]. 또한, 2017년 말까지는 5,000개 이상의 공장을 목. 로 하여 제조업 혁신을 가속화할 예정이라며 인재 양성 계획, 연구 개발 계획 등 정책 방향을 제시하였다[3,5].

이처럼 '제조업 혁신 3.0' 이후 스마트팩토리 구축은 정부의 지원에 힘입어 더욱 가속을 내고 있으며 제조업의 혁신을 꿈꾸는 중소·중견기업의 뜨거운 관심을 받고 있다. 또한, 중소·중견기업의 스마트팩토리 추진 전략과 구축 시 고려사항에 관한 연구도 이루어지고 있다. 그러나 도입 이후의 연구는 찾아보기 힘들며, 주로 성공 사례에 치우쳐있다. 박종식 등(2017)은 스마트팩토리의 개념과 범위, 국내외 동향 등에 대해 연구하여 수요·공급 산

업별 스마트 팩토리 구축 전략과 추진 전략에 대해 제시하였으며[6], 박종필(2017)은 스마트팩토리를 도입한 국내 대·중·소기업의 성공 사례를 분석하여 효율적인 구축 전략을 제시하였으며, 기업의 규모에 따라 구축 방식을 상이하게 진행해야 하며 대기업은 일부 공장을 시범 구축한 후, 전체로 확대하는 전략, 중소기업은 낮은 수준의 스마트팩토리 구축에서 높은 수준으로 업그레이드 하는 전략을 펴야 한다고 제시하였다[7]. 또한, 임명성(2016)은 독일의 제조업 혁신 전략을 살펴본 후, 한국의 제조업 3.0을 성공적으로 추진할 수 있는 전략을 제시하였다. 구호에 그치는 것이 아닌 기업, 연구기관, 학계 등의 협업으로 정부의 지원 하에 제조업 혁신을 추진해야 한다고 제안하였다[8].

스마트팩토리는 해당 시스템의 구축을 통한 제조 효율화 달성, 시스템 유지 보수 과정상에서 발생하는 다양한 경험 축적 등이 이루어져야 한다.[9] 이 시점에서 스마트공장 추진 기업에 대해 점검을 하여 올바른 방향으로 나아가고 있는지 확인할 필요가 있다.

이에 본 연구에서는 스마트팩토리의 기초단계를 적용한 중소·중견기업을 대상으로 설문조사를 하여 스마트팩토리 도입 기업의 현황에 대해 파악하고, 군집분석을 사용하여 특성을 확인하고자 한다. 또한, 특성이 어떠한 형태로 도출이 되는지 알아보기 위해 의사결정나무와 나이브베이즈 분석을 실시해보고자 한다.

2. 이론적 배경

2.1 스마트팩토리

2.1.1 스마트팩토리 개념

스마트팩토리(Smart Factory)란 설비와 물류 자동화를 기반으로 한 공정자동화, 공장자동화, 제품개발, 공급사슬관리, 그리고 기업자원관리 등이 ICT를 사용하여 구현되는 공장을 의미한다. 스마트공장에 대한 다양한 정의가 존재하지만 독일의 인더스트리 4.0이나 미국의 산업인터넷에서 설명하는 내용을 살펴보면, 공장자동화(FA, Factory Automation)가 ICT를 기반으로 하여 발전된 형태로서, IoT, 빅데이터, 클라우드, 사이버물리시스템(Cyber Physical System, CPS), 스마트센서, 3D 프린팅 등 다양한 ICT 기술과 제품의 기획·설계, 생산, 유통·판매 등 제조 전 과정 뿐만이 아니라 기업 경영 전반에

걸쳐 지능적으로 융합된 산업이라고 볼 수 있다[10]. 스마트팩토리의 궁극적인 목적은 생산 시스템을 지능화·유연화·최적화·효율화하여 생산성 향상 및 생산비용 절감을 달성하고 빠르게 변하는 외부적 환경과 고객 요구에 능동적으로 대응할 수 있도록 제조 기업을 구현하는 것이다[11].

스마트팩토리에는 생산설비(시스템)를 기반으로 하는 수직적 통합과 가치사슬을 기반으로 하는 수평적 통합이 존재한다. 수직적 통합은 생산의 효율성을 목표로 한다. 이를 위해 제품 생산에 필요한 다양한 설비에 센서 및 장치를 부착하여 신호를 얻고, PLC (Programmable Logic Controller), HMI(Human Machine Interface) 등의 제어 기술을 통해 설비 제어를 하며, 생산 프로세스 관리를 위해 MES(Manufacturing Execution System), 창고 관리를 위해 WMS(Warehouse Management System)를 거쳐 상단의 ERP(Enterprise Resource Planning)까지 유기적으로 관리하는 개념이다. 수평적 통합은 제품을 사용하는 고객의 요구사항을 도출하기 위한 시장 조사, 제품 기획 단계 후, 고객의 요구사항을 반영한 제품개발 R&D와 공정 설계를 거쳐, 제품 생산 후 고객에게 전달하는 과정까지 모두를 포함한다. 즉, 고객의 요구사항을 시작으로 하여 고객의 필요사항에 맞는 제품을 개발하고 고객에게 전달하기까지의 과정을 관리하는 개념이다[6].

수직·수평적 통합을 유기적으로 구현하기 위하여 다양한 ICT기술이 적용되고 있다. 수평적 통합의 지원 기술에는 제품 설계를 위한 CAD/CAE 등을 포함하는 PLM 솔루션, 시제품 생산을 빠르게 지원하는 3D프린터, 가상과 실제의 연동이 가능한 사이버물리시스템, 제조 프로세스 분석을 위한 공정 시뮬레이션 등이 포함되며, 수직적 통합의 지원 기술에는 생산설비 데이터를 많이 획득하기 위한 스마트센서, IoT기술, 생산 현장의 에너지 절감 기술, 제조 데이터 분석을 위한 제조 빅데이터 기술 등이 포함된다[6].

민관합동 스마트팩토리 추진단에 제시된 스마트공장 수준별 단계를 살펴보면, ①Exel 정도 활용, 시스템을 갖추지 못한 상태인 'ICT 미적용 단계', ②생산 실적 정보를 자동 집계하여 자체흐름을 실시간 파악하고 부분적 관리 시스템을 운영하는 '기초 단계', ③설비 정보를 자동 집계하여 실시간 공장 운영 모니터링과 품질분석을 실시하고 분야별 관리 시스템간 부분적 연계가 이루어진 중간1단계, ④관리 시스템을 통한 설비의 자동제어로 실

시간 생산을 최적화하고 분야별 관리 시스템간 실시간 연동이 이루어지는 중간2단계, ⑤설비, 자재, 시스템이 유무선 네트워크로 연결되어 스스로 판단하는 지능형 설비, 시스템을 통한 자율적 공장 운영이 가능한 '고도화 단계'가 있다.¹⁾

2.1.2 스마트팩토리에서의 빅데이터

제조업의 새로운 혁신 정책으로 스마트팩토리가 추진되면서 설비장치에서 발생하는 다양한 분야의 데이터를 빅데이터 기술과 연결하여 분석하고 있다. 이러한 변화 흐름은 제조회사들의 빅데이터에 대한 관심에서 확인할 수 있다. 포브스(Forbes)에서 연구한 결과, 47%의 제조 회사들이 빅데이터가 회사 성과에 영향을 미친다고 대답하였으며[12], SCM World 조사에서는 미래 제조 환경의 운영과 관리 방법을 바꾸는 데 가장 큰 영향을 미칠 요소에 68%로 빅데이터 분석을 꼽았다. 특히 빅데이터 기술 중 실시간 분석이 가장 큰 영향을 미칠 것으로 보았다. 응답 순위로 확인해보면 1위인 '실시간 공장 가동 효율성 관리 분야'가 57%, 2위인 '실시간 공급망 관리'가 42%로 실시간 관리가 제조업 핵심 활용 분야라 답했다[13].

2.2 데이터마이닝

2.2.1 군집분석

군집분석은 여러 개의 측정치들로 구성된 레코드들로부터 개별 레코드 사이의 거리를 측정하고, 이 거리에 따라 거리가 가까운 레코드끼리 군집화하는 유사한 레코드들의 모임이다[12]. 쉽게 풀이하면, 데이터의 집합 내에서 규칙적인 구조를 발견하고자 하는 목적을 가지고 있으며, 분석 목적에 맞게 각 군집들을 찾아내는 방법을 말한다[15]. 일반적으로 군집화 알고리즘에는 두 가지 유형이 있으며, 각각의 방법론에 따라 여러 가지 구현 알고리즘이 존재한다[16].

계층적 방법은 n개의 레코드를 각각 하나의 군집으로 간주하여 특성이 비슷한 군집끼리 합해나가는 응집 분석(agglomerative analysis)과 전체 레코드를 하나의 군집으로 간주한 후, 특성이 다른 데이터를 분리하는 분할 분석(divisive analysis)이 있다. 비계층적 방법은 군집의 수를 정하지 않고 군집화를 진행하는 방식이며, 미리 몇 개의 군집으로 나눌 것을 예상한 후 군집의 개수를 제공해

1) 스마트공장 수준별 단계, 민관합동스마트팩토리추진단. (<https://www.smart-factory.kr/intro/mainBuz.do>)

야 한다. 비계층적 방법에 가장 많이 쓰는 방법으로는 k-means 분석 방법이 있다. 이 방법은 군집의 수를 정한 후, 각 레코드로부터 미리 정한 k개의 군집의 중심까지 거리를 구하여 가장 가까운 군집에 할당하는 방법이다[14].

2.2.2 의사결정나무

의사결정나무(decision tree) 분석기법은 데이터마이닝 기법 중의 하나로써 의사결정 규칙을 나무구조로 도표화함으로써 관심대상이 되는 집단을 몇 개의 소집단으로 효과적으로 분류하거나 예측하는 기법으로[17], 범주형 반응변수를 사용하는 분류나무와 연속형 반응변수를 사용하는 회귀나무로 나뉜다.

분석 결과는 '조건이 A이고 조건이 B이면 결과 집단 C'라는 형태의 규칙으로 표현이 되므로 분류나 예측을 목적으로 하는 다른 계량적 분석 방법에 비해서 쉽게 이해할 수 있다는 장점을 가진다[18].

나무 모형의 크기를 결정하는 규칙 중 가장 대표적인 것이 정지 규칙과 가지치기 방법이다. 정지규칙은 사용자가 지정한 조건과 일치할 때 나무의 성장을 정지시키는 방법이며, 가지치기는 나무 모형을 우선 크게 만든 후에 불필요한 가지를 제거하여 최적의 나무 모형을 구축해나가는 방법이다[19].

범주형 반응변수를 사용하는 분류나무를 실시할 때, 범주의 수가 많은 경우 분할 가능한 방법의 수가 많아지기 때문에 만약 세 개 이상의 범주로 구성된 범주형 예측 변수를 다룬다면 이를 몇 개의 가변수로 대체하여 사용하여야 한다[14].

의사결정나무분석을 실시하기 위하여 사용되는 알고리즘은 CHAID[20], CART[21], QUEST[22], C4.5[23] 등이 있다.

2.2.3 나이브베이즈

나이브베이즈 분류기(Naive Bayes Classifier)는 머신러닝의 지도학습을 사용한 가장 간단한 기법 중 하나이다. 나이브베이즈 분류는 모형이 단순하며, 계산과정이 간단함에도 분류 성능이 우수하다. 나이브베이즈는 사용되는 데이터의 모든 특성 값이 서로 독립임을 가정하며, 분류를 위해 베이즈 정리(Bayes's Theorem)를 기본으로 사용한다[24].

나이브베이즈는 목표가 확률을 추정하는 것일 때는

매우 편향된 결과를 제공하지만, 레코드의 분류나 순위를 구하는 것이 목표일 때 좋은 성능을 얻을 수 있는 방법이다[25].

베이즈 정리를 사용하는 이유는 조건부 확률을 구할 경우 베이즈 룰을 이용한다면 쉽게 값을 구할 수 있기 때문이다. d를 입력한 값, 분류할 클래스를 c라고 표현할 때, 베이즈 룰을 적용하면 (1)과 같이 나타낼 수 있다.

$$P(c|d) = \frac{P(d|c)P(c)}{P(d)} \dots \dots \dots (1)$$

나이브베이즈는 모든 분류에 대한 확률을 계산한 후 가장 높은 확률을 가지는 분류를 정한다. 최종적으로 나이브베이즈 식은 (2)와 같다.

$$\begin{aligned} C_{MAX} &= \operatorname{argmax} P(c|d) \\ &= \operatorname{argmax} \frac{P(d|c)P(c)}{P(d)} \dots \dots \dots (2) \\ &= \operatorname{argmax} (d|c)P(c) \end{aligned}$$

3. 기초 분석

3.1 데이터 설명

본 연구는 스마트팩토리 기초단계에 적용한 기업을 대상으로 연구를 실시하였다. 민관합동 스마트팩토리 추진단에서는 ICT 미적용 기업을 대상으로 스마트팩토리 도입을 지원하고 있으므로 지원을 통해 기초단계에 적용한 기업이 설문 대상으로 선정되었다. 설문지는 스마트팩토리 추진위원회의 전문가 의견을 토대로 작성하였으며, 조사 기간은 2017년 1월 2일부터 2017년 1월 26일까지 실시하였다. 조사 방법으로는 스마트팩토리 담당자의 의견을 구글 설문지를 통해 설문 응답을 받았다. 25개의 변수에 대해 응답을 받아 총 140건의 설문지를 회수하였고, 이 중 응답 상태가 불량한 15건을 제외한 125건을 분석에 사용하였다.

3.2 전체 데이터 대상의 군집분석

먼저, 수집한 설문데이터의 특성을 파악하기 위해 데이터 마이닝 툴인 WEKA 3.6을 사용하여 군집분석을 실

시하였다. 125개의 데이터에 대해 군집분석을 실시하였으며, 사용된 변수는 25개이다.

군집의 수를 27개로 변경하며 분석을 수행한 결과 뚜렷한 군집의 특징을 찾기 어려웠다. 그 중 4개의 군집으로 분류를 하였을 때 다소 유의미한 특징이 도출되었다. 4개의 군집 분류 결과, 만족도 변수를 기준으로 만족도가 낮은 두 군집과 만족도가 높은 두 군집으로 분류되었다.

만족도가 1점, 2점, 5점인 기업의 수가 적고 만족도가 3점, 4점인 기업의 수가 많기 때문에 만족도가 낮은 군집과 만족도가 높은 그룹의 만족도 차이는 크지 않다. 그러나 평균 만족도를 기준으로 만족도가 낮은 그룹과 만족도가 높은 그룹으로 나누었을 때 군집분석 결과 차이가 분명하기에 만족도가 낮은 그룹과 만족도가 높은 그룹 기준으로 나누어 분석을 하였다. 평균 만족도 3.44를 기준으로 만족도가 낮은 군집은 Cluster0과 Cluster2이고, 만족도가 높은 군집은 Cluster1과 Cluster3이다.

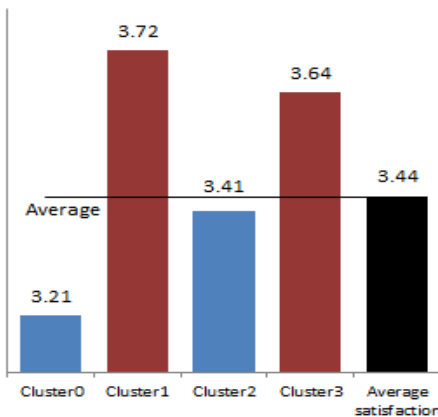


Fig. 1. Satisfaction by cluster

3.2.1 만족도가 낮은 기업

Cluster0은 47개 기업으로, 만족도는 3.21이다. 5년 이내에 중간 1단계의 설비를 갖추는 것이 적정수준이라고 생각하며 스마트팩토리 도입 후 매출 증가율이 5%이하인 그룹이다.

Cluster2는 31개 기업으로 만족도는 3.41이다. 스마트팩토리 도입 후 매출 증가율이 5%이하이며 스마트팩토리 고도화를 10년 이내에 추진할 계획이 있는 그룹이다.

3.2.2 만족도가 높은 기업

Cluster1은 22개 기업으로 만족도는 3.72이다. 5년 이

내 최종단계의 설비를 갖추는 것이 적정수준이라 생각하며, 스마트팩토리 도입 후 매출 증가율이 10%이하인 그룹이다.

Cluster3은 25개 기업으로 만족도는 3.64이다. 스마트팩토리 도입 후, 매출 증가율이 10%이하, 불량률 개선이 20%이하인 그룹이다.

만족도가 낮은 그룹인 Cluster0, Cluster2와 만족도가 높은 그룹인 Cluster1, Cluster3을 비교하여보면, 만족도가 낮은 그룹은 만족도가 높은 그룹에 비해 매출 증가율과 불량률 개선이 낮으며 비교적 일정 기간 내의 도달하고자 하는 목표가 낮은 것을 확인할 수 있다.

이 결과를 바탕으로 본 논문은 만족도의 평균이 3.4점이기 때문에 3점은 만족도가 낮은 것으로 분류하였으며, 만족도가 4-5점인 기업과 만족도가 1-3인 기업으로 나누어 각각 군집분석을 실시하여 특성을 알아보하고자 한다. 또한, 만족도가 높은 기업을 중점으로 의사결정나무와 나이브베이지를 실시하여 어떠한 특성으로 만족도가 도출될지 예측해보고자 하며, 두 예측 방법 중 어떠한 분석 방법이 더 정확한지 확인해보고자 한다.

4. 심도 분석

4.1 일반적 특성

응답 기업들의 일반적 특성을 확인해보면, 만족도가 높은 기업은 주로 충청도와 경기도 지역에 많으며, 업종은 기타업종 58%, 전자부품 14%, 표면처리 8% 순으로 많다. 임직원은 20명 이상~100명 미만이 절반 이상을 차지하며 매출은 50억 미만과 400억 이상이 23%로 가장 많다.

만족도가 낮은 기업은 주로 경상도, 경기도 지역에 많고 업종은 기타업종 52%, 주조, 금형이 10%로 많다. 임직원은 만족도가 높은 기업과 마찬가지로 20명 이상~100명 미만이 절반 이상을 차지하며, 매출은 400억 이상이 25%로 가장 많다.

4.2 군집 분석

전체 대상의 군집분석 결과, 평균 만족도(3.44점)를 기준으로 만족도가 낮은 집단과 만족도가 높은 집단으로 나누었다. 만족도에 따라 1-3점을 만족도가 낮은 집단, 4-5점을 만족도가 높은 집단으로 분류하였고, 그 결과 만

족도가 낮은 집단은 60개, 만족도가 높은 집단은 65개로 분류되었다. 분류 후, 데이터 마이닝 툴인 WEKA 3.6을 사용하여 만족도가 높은 집단과 만족도가 낮은 집단에 대하여 각각 군집 분석을 실시하였다. 사용된 변수는 ‘만족도’ 변수를 제외한 24개 변수이다.

4.2.1 만족도가 낮은 집단

만족도가 낮은 집단에 대해 군집분석을 실시한 결과 2개의 군집으로 구분이 되었으며 Cluster 0은 26개, Cluster 1은 34개로 분류되었다.

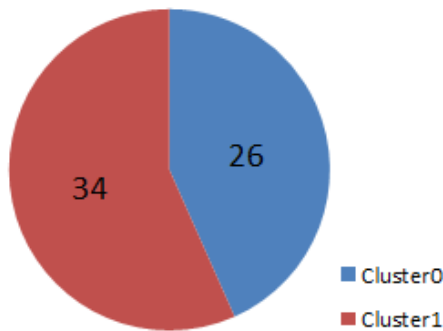


Fig. 2. Cluster analysis of low satisfaction group

가. cluster 0 (불량률 개선을 희망하는 군집)

cluster 0은 MES를 적용하였으며, 불량률 개선이 평균 8%로 낮은 군집이다. 이 군집은 생산관리 시스템인 MES를 도입하여 최적의 생산 활동을 기대하였지만, 불량률에서 큰 개선을 보이지 않아 그로 인해 스마트팩토리 도입에 대한 만족도가 낮음을 볼 수 있다. 이유를 상세히 살펴보니, 이 군집의 기업들은 측정설비연계와 생산설비연계가 매우 낮아 자동화 시스템을 통해 서버에 데이터를 전송하는 양이 적으므로 설비별 데이터를 활용하여 불량률, 생산성 등의 성과를 거둘 수 없었던 것으로 생각할 수 있다.

나. cluster 1 (품질 개선을 희망하는 군집)

cluster 1은 ERP를 적용하였으며, 품질 향상에 대한 기대가 큰 군집이다. ERP 시스템은 기업 내 생산, 물류, 재무, 회계, 영업/구매, 재고 등 경영활동에 관한 프로세스들을 통합 연계하여 업무 프로세스를 개선해주는 시스템이기 때문에 직접적인 품질 향상을 가져다주지 않는다. 따라서 품질 향상에 대한 기대로 스마트팩토리를 추진하

였으나, ERP 시스템으로 품질향상을 이루지 못한 만족도가 낮은 군집으로 볼 수 있다.

4.2.2 만족도가 높은 집단

만족도가 높은 집단에 대해 군집분석을 실시한 결과 3개의 군집으로 구분이 되었으며 Cluster 0은 11개, Cluster 1은 23개, Cluster 2는 31개로 분류되었다. cluster0은 생산성 개선, 매출증가와 불량률 개선이 가장 낮은 군집(생산성 개선:12%, 매출 증가:7%, 불량률 개선:10%)이었으며, cluster2는 생산성 개선, 매출증가와 불량률 개선이 가장 높은 군집(생산성 개선:26%, 매출 증가:18%, 불량률 개선:22%)이었다. cluster0의 생산성 개선, 매출증가와 불량률 개선은 만족도가 높은 기업의 군집 중 가장 낮지만, 만족도가 낮은 군집의 두 군집보다는 높은 것을 볼 수 있었다.

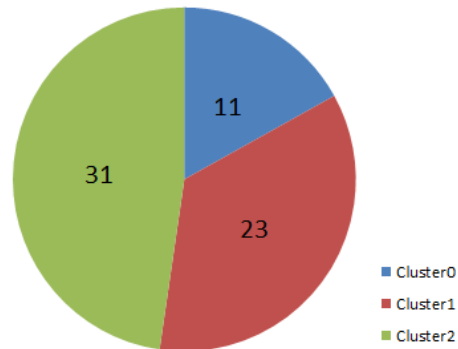


Fig. 3. Cluster analysis of high satisfaction group

가. cluster 0 (추진 의지가 높은 군집)

cluster 0은 기초단계에 체류를 원하지 않으며, 2차 정부 지원이 불가하더라도 관계없이 추진하겠다는 군집이다. 이 군집은 두 항목에서 일관성 있게 더 높은 단계인 중간 1단계로 나아가길 원하는 것으로 보아 추진 의지가 높은 군집으로 볼 수 있다.

나. cluster 1 (정부 지원을 기대하는 군집)

cluster 1은 MES를 적용하였고, 중간 1단계 추진시 정부 지원이 가장 중요하며 기초단계에 체류하는 이유가 투자비용인 군집이다. 이 군집은 MES도입으로 생산성, 매출 증가와 불량률 개선에서 성과를 이루었으며 중간 1단계로 나아가고 싶지만, 투자비용 때문에 현재 기초단

계에서 머물고 있는 군집이다. 이 군집은 정부 지원을 통해 중간 1로의 추진을 기대하고 있는 것으로 보인다.

③ cluster 2 (높은 성과를 이룬 군집)

cluster 2는 생산 설비 연계 비율과 측정 설비 연계 비율이 높으며 생산성 개선, 매출 증가, 불량률 개선이 가장 높은 군집이다. 이 집단은 설비 연계 비율이 가장 높아 생산성 개선, 매출 증가, 불량률 개선에서 큰 성과를 이룬 집단이므로 만족도가 높은 것으로 보인다.

4.3 의사결정나무

4.3.1 범주의 조정

세 개 이상의 범주로 구성된 범주형 예측변수를 사용할 때는 변수를 조정하여 의사결정나무의 분류를 실시해야 한다. 따라서 현재 5개의 범주로 구성된 예측변수의 수를 조정하여 의사결정나무를 실시하였다. 가운데 범주인 3번 범주를 기준치로 하여 기준치 미만과 기준치 이상인 두 가지 범주로 나누었다. 그러나 0의 개념인 계획 없음, 없음, 전혀 안 됨, 해당 없음, 적용 안 함이 포함된 변수의 범주는 계획 없음, 없음, 전혀 안 됨, 해당 없음, 적용 안 함을 포함하여 3개의 범주로 분류하였다.

4.3.2 의사결정나무 분석

의사결정나무를 사용하여 분석을 실시할 때, 업종과 지역 변수는 제외하였다. 업종과 지역은 해당 기업의 현황을 파악하는 자료로 사용되었고 범주를 조정하지 않았기 때문에 분석에 사용하지 않았다. 만족도를 목표변수로 하여 의사결정나무 분석을 실시한 결과 72%의 예측 정확성을 보였다.

분석 결과는 다음과 같다.

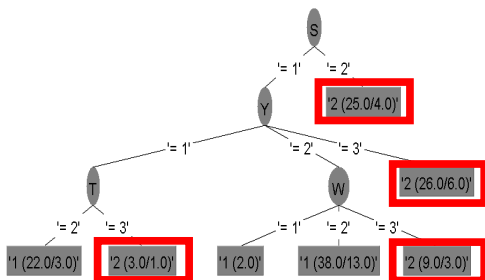


Fig. 4. Decision tree analysis results.

가. 생산성개선(S)정도가 높은 기업은 만족도가 높다.
나. 생산성개선(S)정도가 낮고 스마트팩토리 추진속도(Y)가 빠르다고 생각하는 기업은 만족도가 높다.

다. 생산성개선(S)정도가 낮고, 스마트팩토리 추진속도(Y)가 느리며 스마트팩토리 적용 기간(T)이 2년 이상인 기업은 만족도가 높다.

라. 생산성개선(S)정도가 낮고, 스마트팩토리 추진속도(Y)가 보통이며 매년 투자액(W)이 1억 이상인 기업은 만족도가 높다.

4.4 나이브 베이즈

마찬가지로 나이브 베이즈 분석을 실시할 때 해석 결과를 용이하게 하기 위해 범주를 조정한 데이터를 사용하여 분석을 실시하였으며, 업종과 지역 변수는 제외한 후 분석했다. 만족도를 목표변수로 하여 나이브베이즈를 실시한 결과 66%의 예측 정확성을 보였다.

전체 데이터에서 만족도가 낮은 기업이 48%, 만족도가 높은 기업이 52%를 차지하며, 구체적인 해석은 다음과 같다.

가. 3년 후 적정수준(F)에서 만족도가 낮은 그룹은 중간 1단계(39건), 만족도가 높은 그룹은 중간 2단계(27건)의 수가 많다.

나. 적용시스템(G)은 만족도가 낮은 그룹은 ERP(31건), 만족도가 높은 그룹은 MES(35건)가 많다.

다. 중간 1단계 추진 시 중요사항(J)에서는 만족도가 낮은 그룹은 운영안정화(21건), 만족도가 높은 그룹은 적극적 활용(25건)이 높다.

라. 측정설비연계(L)와 생산설비연계(M)는 만족도에 관계없이 50%이하가 가장 많지만, 만족도가 높은 그룹은 만족도가 낮은 그룹에 비해 50%이상의 건수(측정설비연계 14건, 생산설비연계 22건)가 높다.

마. 5년 이내 적정수준(N)은 만족도가 낮은 그룹은 중간 1단계(31건), 만족도가 높은 그룹은 중간 2단계(30건)의 수가 많다.

바. 기초단계체류 이유는 만족도에 관계없이 투자비용이었지만, 그 다음으로는 만족도가 높은 그룹은 체류를 원하지 않음(15건)이고, 만족도가 낮은 그룹은 전문가부족(14건)으로 나타남.

Table 1. Table title

Variable	Satisfaction		Variable	Satisfaction		Variable	Satisfaction		Variable	Satisfaction	
	low (0.48)	high (0.52)		low (0.48)	high (0.52)		low (0.48)	high (0.52)		low (0.48)	high (0.52)
C(Sales)			H(Intermediate 1 stage Promotion)			N(Appropriate level within 5 years)			S(Productivity Improvement)		
1	36.0	39.0	1	9.0	4.0	1	4.0	5.0	1	57.0	45.0
2	24.5	28.0	2	43.0	54.0	2	31.0	25.0	2	5.0	22.0
total	46.0	67.0	3	11.0	10.0	3	22.0	30.0	total	62.0	67.0
D(Employees)			total	63.0	68.0	4	7.0	9.0	T(Application period)		
1	42.0	49.0	J(Intermediate 1 stage Requirements)			total	64.0	69.0	2	51.0	51.0
2	20.0	18.0	1	21.0	15.0	O(Stay at the basic stage)			3	11.0	16.0
total	62.0	67.0	2	2.0	2.0	1	8.0	15.0	total	62.0	67.0
E(Intermediate 1 stage biggest difficulty)			3	14.0	10.0	2	35.0	37.0	U(Increase in sales)		
2	41.0	54.0	4	16.0	25.0	3	14.0	9.0	1	58.0	55.0
3	9.0	7.0	5	12.0	18.0	4	3.0	3.0	2	4.0	12.0
4	8.0	5.0	total	65.0	70.0	5	5.0	6.0	total	62.0	67.0
5	6.0	3.0	K(Director)			total	65.0	70.0	V(Promotion of advanced stage)		
total	64.0	69.0	1	6.0	7.0	P(No government second support)			1	28.0	23.0
F(Difficult level after 3 years)			2	45.0	46.0	1	3.0	4.0	2	20.0	24.0
1	1.0	3.0	3	6.0	7.0	2	7.0	7.0	3	15.0	21.0
2	8.0	9.0	4	4.0	7.0	3	3.0	1.0	total	63.0	68.0
3	39.0	22.0	5	4.0	3.0	4	43.0	46.0	W(Investment)		
4	12.0	27.0	total	65.0	70.0	5	9.0	12.0	1	5.0	2.0
5	5.0	9.0	L(Connection of measuring equipment)			total	65.0	70.0	2	47.0	45.0
total	65.0	70.0	1	22.0	16.0	Q(Company History)			3	11.0	21.0
G(System)			2	37.0	38.0	1	18.0	24.0	total	63.0	68.0
1	28.0	35.0	3	4.0	14.0	2	44.0	43.0	X(Improved defect rate)		
2	31.0	30.0	total	63.0	68.0	total	62.0	67.0	1	57.0	49.0
3	2.0	2.0	M(Connection of production facilities)			R(Motivation)			2	5.0	18.0
4	2.0	2.0	1	18.0	13.0	1	15.0	18.0	total	62.0	67.0
5	2.0	1.0	2	32.0	33.0	2	19.0	18.0	Y(Propulsion speed)		
total	65.0	70.0	3	13.0	22.0	3	7.0	9.0	1	21.0	9.0
-			total	63.0	68.0	4	23.0	23.0	2	34.0	26.0
						5	1.0	2.0	3	8.0	33.0
			total	65.0	70.0	total	63.0	68.0			

사. 생산성개선(S)과 불량률개선(X)은 만족도에 상관 없이 20%이하인 비율이 가장 높지만, 만족도가 높은 기업은 만족도가 낮은 기업에 비해 20%이상인 기업(생산성 개선: 22건, 불량률 개선: 18건)이 상당히 많다.

아. 고도화 추진시기(V)는 만족도가 낮은 기업은 고도화를 도입하지 않거나 10년 이후인 기업(28건)이 가장 많고, 만족도가 높은 기업은 고도화 도입을 5년 이내로 답한 기업(24건)이 가장 많다.

자. 매년 투자액(W)은 만족도에 관계없이 1억 이하가 가장 많지만, 만족도가 높은 기업은 1억 이상이라고 답한 기업(21건)이 만족도가 낮은 기업에 비해 상당히 많다.

차. 스마트팩토리 추진속도(Y)는 만족도가 낮은 기업은 동종기업대비 늦거나(21건) 중간(34건)이라고

답한 기업이 많고, 만족도가 높은 기업은 동종기업 대비 중간(26건)이나 빠르다(33건)고 답한 기업이 많다.

5. 연구 결과 논의

본 연구에서는 스마트팩토리의 기초단계에 있는 중소·중견기업을 대상으로 설문조사를 실시하여 스마트팩토리 도입 기업의 현황에 대해 파악하고, 군집분석을 사용하여 특성을 확인하였다. 또한, 특성이 어떠한 형태로 도출이 되는지 알아보기 위해 의사결정나무와 나이브베이즈 분석을 실시하였다.

먼저, 전체 데이터의 특성을 확인하기 위해 전체 기업을 대상으로 군집분석을 실시하여 특징을 살펴보았다.

분석 결과, 네 개의 군집으로 분류하였을 때, 유의미한 결과를 볼 수 있었으며 평균 만족도(3.44점)를 기준으로 두 개씩 나누는 것을 볼 수 있었다. 만족도가 낮은 군집은 만족도가 높은 그룹에 비해 매출 증가율과 불량률 개선이 낮으며 비교적 일정 기간 내의 도달하고자 하는 목표가 낮은 것을 확인할 수 있었다.

전체 데이터 대상의 군집분석을 실시한 후, 만족도가 높은 그룹과 만족도가 낮은 그룹의 개별 특징을 확인하기 위해 데이터를 만족도별로 나누어 군집분석을 실시하였다.

만족도가 낮은 그룹은 불량률 개선을 희망하는 군집(MES를 적용하고 불량률 개선이 8%로 낮은 군집)과 품질 개선을 희망하는 군집(ERP를 적용하고 품질향상에 대한 기대가 큰 군집)으로, 총 2개의 군집으로 나뉘었다. 만족도가 높은 그룹은 추진의지가 높은 군집(기초단계에 체류를 원하지 않고 2차 정부 지원이 없더라도 관계없이 추진하는 군집), 정부지원을 기대하는 군집(MES를 적용하고 중간 1 추진시 정부 지원이 가장 중요하며 기초단계 체류 이유가 투자비용인 군집), 그리고 높은 성과를 이룬 군집(생산, 측정 설비 연계 비율 가장 높으며 생산성 개선, 매출 증가, 불량률 개선이 가장 높은 군집)으로, 총 3개의 군집으로 나뉘었다.

만족도별 군집분석을 통해 각각의 특징을 확인한 후, 의사결정나무와 나이브베이지를 실시하기 위해 데이터를 다시 통합하였으며, 결과 해석을 용이하게 하기 위해 변수의 응답 범주를 조정하였다. 기존 5개의 응답 범주를 2개 혹은 3개로 축소하여 해석하기 편하게 변경하였다.

의사결정나무 분석 결과, 72%의 정확성을 보였고 ‘생산성 개선 정도’를 기준으로 가지가 나뉘었다. ‘생산성 개선 정도’가 높으면 만족도가 높고, ‘생산성 개선 정도’가 낮고 ‘동종 기업 대비 스마트팩토리 추진 속도’가 빠르다고 생각하면 만족도가 높았다. 이 결과 외에도 만족도가 높은 기업의 특징들을 확인할 수 있었다.

나이브베이지 분석 결과, 66%의 정확성을 보였다. 결과를 해석하여보면 ‘3년 후 적정수준’은 만족도가 낮은 그룹보다 만족도가 높은 그룹에서 더 높은 단계를 희망하는 것을 볼 수 있었다. ‘적용시스템’은 만족도가 낮은 그룹은 ERP, 만족도가 높은 그룹은 MES가 많은 것을 볼 수 있었다. 이 결과 외에도 다양한 결과를 도출할 수 있었다.

예측 방법인 의사결정나무와 나이브베이지 분석을 실

시한 결과, 정확성은 의사결정나무가 72%로 나이브베이지(66%)보다 높은 것을 볼 수 있었다. 그러나 의사결정나무의 결과는 나이브베이지와 비교하여 보았을 때 도출할 수 있는 의미 정보가 적은 것으로 확인되었다. 분석의 정확도는 낮지만 다양한 결과를 얻기 위해선 나이브베이지를 활용하는 것이 나은 것으로 확인되었다.

데이터마이닝 분석 결과를 통합하여 의미를 해석하였다. 생산성 개선, 불량률 개선, 측정설비연계, 생산설비연계, 매년 투자액, 스마트팩토리 추진속도 등이 만족도에 영향을 주는 요인인 것으로 확인할 수 있었다.

6. 결론

본 연구에서는 군집분석을 통해 만족도별 기업의 특성을 확인한 후, 두 가지 예측 분석 방법인 의사결정나무와 나이브베이지를 비교하여 만족도가 어떠한 특성으로 도출되는지 확인해보았다. 이 과정에 예측의 정확성과 의미 해석의 다양성은 서로 상반성이 없는 것을 확인하였으며, 의사결정나무보다 나이브베이지의 정확성이 다소 낮더라도 다양한 의미 해석이 가능하다는 것을 볼 수 있었다. 본 연구는 스마트팩토리 도입 예정인 기업에게 어떠한 요소에 중점을 두고 구축을 해야 하는지 참고자료로 사용될 수 있도록 실제 기업 데이터를 바탕으로 만족도별 특징을 제시한다는 점에서 의의가 있다. 또한, 현재 증가하는 스마트팩토리에 관한 연구에 참고할 수 있도록 데이터마이닝 기법을 적용하여 기업별 특성을 확인하였다는 데에 의의가 있다.

본 연구에는 몇 가지 한계가 있다. 먼저 데이터의 수가 125개로 매우 적으며 이는 과적합되어 다소 부정확한 결과를 도출할 수 있다. 데이터 수를 늘린다면 의사결정나무에서도 다양한 결과를 도출할 수 있을 것이며, 나이브베이지의 정확성을 높일 수 있을 것이다. 또한, 스마트팩토리 기초단계를 도입한 기업 대상으로 설문조사 실시되었기 때문에, 기초단계가 아닌 중간 1, 중간 2, 고도화 단계에서는 새로운 연구가 이루어져야 할 것이다. 마지막으로, 설문 답변 범주를 홀수가 아닌 짝수로 변경한 후 ‘보통’이나 ‘중간’에 대한 답변을 제외하여 설문을 한다면 좀 더 정확한 결과를 도출할 수 있을 것이다.

REFERENCES

- [1] Deloitte. (2015). *The US recovers the top position in the global manufacturing competitiveness in 2020*, Deloitte Korea.
<https://www2.deloitte.com/kr/ko/footerlinks/pressrelease/page/2015/press-release-20151211.html>
- [2] Y. J. JO. (2015). *Possibility of smart factory as a plan for advanced domestic manufacturing*. KDB Bank.2015.8.21
https://rd.kdb.co.kr/er/wcms.do?actionId=ADERERERWCE03&contentPage=/er/er/ERER27100012_01RS.jsp&menuId=ERERER0013&cid=19772
- [3] C. W. Lee & Y. B. Jang. (2017). *Leading the Fourth Industrial Revolution by Building 30,000 Smart Factories by 2025*. KOSF(Korea Smart Factory Foundation).
<https://www.smart-factory.kr/datum/popup/datumDetail.do?dboardNo=121>
- [4] M. K. Jung. (2016). *Smart factory, check performance and revisit the past two years*. KOSF(Korea Smart Factory Foundation).
<http://www.smart-factory.kr/Service/Notice/appl/ReportView.asp>
- [5] H. S. Lee. (2017). *4th Industrial Revolution Leading Smart Factory, 5,000 Spreads by Year*. KOSF(Korea Smart Factory Foundation).
<http://www.smart-factory.kr/Service/Notice/appl/ReportView.asp>
- [6] J. S. Park & K. S. Kang. (2017). Strategies of smart factory building and Application of small & medium-sized manufacturing enterprises. *Korea Safety Management & Science*, 19(1), 227-236.
- [7] J. P. Park (2017). Analysis on Success Cases of Smart Factory in Korea: Leveraging from Large, Medium, and Small Size Enterprises, *The Korea Society of Digital Policy and Management*, 15(5), 107-115.
- [8] M. S. Lim (2016). (The)Convergence between Manufacturing and ICT : The Exploring Strategies for Manufacturing version 3.0 in Korea, *The Korea Society of Digital Policy and Management*, 14(3), 219-226.
- [9] T. S. Jeong. (2016). The Suggestion for Successful Factory Converging Automation by Reviewing Smart Factories in German, *Journal of the Korea Convergence Society*, 7(1), 189-196.
- [10] J. Hoh & C. Y. Jung (2017). Convergence-based Smart Factory Security Threats and Response Trends, *Journal of the Korea Convergence Society*, 8(11), 29-35.
- [11] C. S. Seo (2016). *Study on Small Business Increased Productivity via Smart Factory*. Master dissertation. Busan National University, Busan.
- [12] C. Louis. (2015). *Data Analytics, Mobile Technologies And Robotics Defining The Future Of Digital Factories*, Forbes.
<https://www.forbes.com/sites/louiscolombus/2015/02/15/big-data-analytics-mobile-technologies-and-robotics-defining-the-future-of-digital-factories/#137e08fe7e9d>
- [13] S. R. Jo. (2015). *SmartFactory. June*, Industry Soutlution
http://www.google.co.kr/url?sa=t&rct=j&q=&esrc=s&source=web&cd=2&ved=0ahUKEwjz3eit1sjZAhVHIzQKHapKBOoQFggsMAE&url=http%3A%2F%2Ffile10.uf.tisto ry.com%2Fattach%2F27488035557A86CB2C79A7&usg=AOvVaw3vNkBlgv3Mff51GivZs_4K
- [14] Shmueli, G., R. P. Nitin & C. B. Peter. (2012). *Data Mining for Business Intelligence*. Seoul : E&B Plus.
- [15] J. G. Jo & S. H. Choi (2016). Firm's Market Value Trends after Information Security Management System(ISMS) Certification acquisition, *Journal of the Korea Convergence Society*, 7(6), 237-247.
- [16] He, Q. (1999). *A Review of Clustering Algorithms as Applied in IR*. Technical Report UIUCLIS-1999/6+IRG, University of Illinois at Urbana-Champaign.
- [17] J. S. Bae. (2014). A Study on Priority of Determinants of Career Decision Level in High School Students Based on Decision Tree Analysis. *The Korean Society for the Study fo Career Education*, 28(4), 79-105.
- [18] Nguyen, T. D., T. B. Ho & H. Shimodaira. (2001). A Scalable Algorithm for Rule Post-pruning of Large Decision Trees. *5th Pacific-Asia Conference on Knowledge Discovery and Data Mining*, 467-476.
- [19] N. Y. Park, J. I. Kim & Y. G. Jung. (2013). Breast Cancer Diagnosis using Naive Bayes Analysis Techniques. *The society of Service Science*, 3(1), 87-93.
- [20] Kass, G. (1980). An exploratory technique for investigation large quantities of categorical data. *Applied Statistics*, 29, 119-129.
- [21] Breiman, L., J. H. Friedman, R. A. Olshen, & C. J. Stone. (1984). *Classification and regression tress*, Wadsworth.
- [22] Loh, W. & Y. Shih.(1997). Split selection methods for classification trees. *Statistica Sinica*, 7, 815-840.
- [23] Quinlan, J. R.(1993). C4.5 Programs for machine learning, *Morgan Kaufmann, San Mateo*.
- [24] H. R. Jeong, H. H. Kim, S. M. Park, K. H. Kim & I. S. Yun. (2017) Prediction of Severities of Rental Car Traffic Accidents using Naive Bayes Big Data Classifier. *Korea Inst. Intelligenct Transportation System*, 2017(4), 411-414.
- [25] K. Larsen. (2005). Generalized Naive Bayes Classifiers. *SIGKDD Explorations*, 7(1), 76-81.

오 정 윤(Oh, Jeong Yoon)

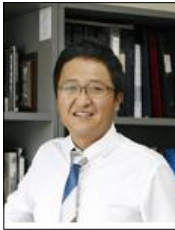
[학생회원]



- 2016년 2월 : 충북대학교 경영정보학과(경영학사)
- 2018년 2월 : 충북대학교 경영정보학과(경영석사)
- 관심분야 : 데이터마이닝, 빅데이터
- E-Mail : jyoony518@naver.com

최 상 현(Sang-Hyun Choi)

[정회원]



- 1998년 2월 : KAIST 경영정보공학 (공학박사)
- 2002년 10월 ~ 2011년 8월 : 경상대학교 산업시스템공학부 부교수
- 2011년 9월 ~ 현재: 충북대학교 경영정보학과 교수
- 2013년 3월 ~ 현재 : 한국빅데이터서비스학회 부회장
- 관심분야 : 빅데이터, 프로세스마이닝, 데이터마이닝, 의사결정지원시스템
- E-Mail : chois@cbnu.ac.kr