

# 빅데이터를 활용한 복지정책 시각화분석 -충청도 중심으로-

김대유<sup>1</sup>, 나원식<sup>2\*</sup>

<sup>1</sup>남서울대학교 빅데이터산업보안학과 조교수, <sup>2</sup>남서울대학교 컴퓨터소프트웨어학과 교수

## Welfare Policy Visualization Analysis using Big Data -Chungcheong-

Dae-Yu Kim<sup>1</sup>, Won-Shik Na<sup>2\*</sup>

<sup>1</sup>Assistant professor, Dept. of Bigdata Industrial Security, Namseoul University

<sup>2</sup>Professor, Dept. of Computer Science, Namseoul University

**요약** 본 연구의 목적은 4차산업혁명 시대의 빅데이터 분석 기술을 활용한 충청도 복지정책 변화와 중요성을 분석하고 사회적 약자를 포함한 모든 세대의 안정적 복지정책을 제안하였다. 충청도 정책 관련 빅데이터를 파이선으로 코딩하여 시각화분석 결과를 토대로 안정적인 정부 정책을 제안한다. 연구 결과 충청도 정부 정책의 키워드는 지역, 사회, 정부 및 지원, 교육, 여성 등의 순으로 확인되었으며, 지역 건강정책과 사회 복지 향상을 중심으로 복지정책을 강화해야 한다. 향후 연구 방향은 해외사례를 비교하고, 전국적인 복지정책의 안정적인 영향에 관한 정책 제안이 필요할 것이다.

**주제어** : 빅데이터, 파이선코딩, 복지정책, 시각화분석, 언론

**Abstract** The purpose of this study is to analyze the changes and importance of welfare policies in Chungcheong Province using big data analysis technology in the era of the Fourth Industrial Revolution, and to propose stable welfare policies for all generations, including the socially underprivileged. Chungcheong-do policy-related big data is coded in Python, and stable government policies are proposed based on the results of visualization analysis. As a result of the study, the keywords of Chungcheong-do government policy were confirmed in the order of region, society, government and support, education, and women, and welfare policy should be strengthened with a focus on improving local health policy and social welfare. For future research direction, it will be necessary to compare overseas cases and make policy proposals on the stable impact of national welfare policies.

**Key Words** : Big Data, Python Coding, Welfare Policy, Visualization Analysis, Media

### 1. 서론

정보와 통신 등에 기술 향상으로 기업 간 경계가 모호해지면서 융합 현상이 새로운 기술혁신의 패러다임으로 떠오르고 있다. 인간의 생각을 대체하는 인공지능, 대용량의 빅데이터, 보안을 중시하는 블록체인, 센서 기

술이 포함된 산업용 사물인터넷(IIOT)등 ICT 기술 기반의 디지털 기술로(Digital Transformation) 대표되는 4차산업혁명이 확산하고 이러한 4차산업혁명 기술 발전이 발전함에 따른 산업체에서는 여러 가지 새로운 사업 기회를 맞이할 것으로 예측 및 국내외시장 확장, 관련·비관련 여러 가지 사업화를 위한 합병, 기업 취득 등 신

\*Corresponding Author : Won-Shik Na(winner@nsu.ac.kr)

구사업 투자(직접, 간접), 새로운 기술 관련 투자, 파괴적 기술 역량 보유 등 안정적인 성장을 위한 여러 가지 전략적 의사결정을 하고 있다[1].

제4차 산업혁명의 시작과 알림은 2016년 46회차의 다보스포럼인 세계경제포럼에서 진행되었다. 비슷한 혁신 전체가 준비 단계로서 화학적이지 않은 디지털 변화(이전의 아날로그), 물리적인 분야의 경우 안전과 변화가 서로 개조 사하게 섹처적인 상황에서 기술적 향상의 출발점에 가까워졌다고 선언하였다. IoT 중심을 스마트한 농업 시스템 중 설비가 융합된 과수에 이용하는 4차 산업기술은 IOT, 빅데이터 활용, 드론 기술, 로봇, 빅데이터 등이 있다[2].

본 연구는 4차산업시대의 빅데이터 시각화분석 기술을 활용한 충청도 지역의 정책 관련 빅데이터를 파이선으로 코딩하여 분석 결과를 토대로 안정적인 정부 정책을 제안하는 것이 주목적이다.

언론진흥재단에서 비정형 빅데이터를 파이선으로 코딩하여 빅데이터 클라우드 워드로 시각화하고 분석 결과를 토대로 지역 건강정책, 사회 복지 향상, 경제적 안정 등을 중심으로 복지정책을 제안한다.

## 2. 이론적 배경

### 2.1 빅데이터 이론

빅데이터의 전체 분야에서 Database Manage System으로 데이터의 보관 분석 최적의 관리 범위를 초과하는 상당히 큰 데이터이다. 최근 정형, 비정형데이터로부터 가치를 활성화하여 의의를 각인하기 위한 기술로서 빅데이터 분석 기법을 활용하고 있다. 빅데이터는 이전의 데이터와 다르게 흔히 데이터를 확보하고 정형화하여 혁신하는 것만 아니라 큰 데이터를 융합하여 한번에 작은 결과를 넘어서 파괴적 혁신 발견 및 예측을 가능케 한다. 우리가 여러 가지 데이터들을 모으고 가공하여 통합적인 데이터들을 활용하여 결론을 예측하는 것과 다르게 빅데이터는 미래에 발생할 결론을 예측한다[3].

여러 가지 조율의 어마한 양의 데이터로부터 선택적으로 필요한 부분을 개발하고 데이터의 신속 안전 정확하게 데이터를 확보하여, 분석하도록 확립된 4차산업혁명시대의 기술로서 보통 테라바이트보다 큰 크기를 가지고 있으며, 다양한 비정형데이터를 포함한 데이터 만들어지고, 유통 판매가 아주 빠르게 분 단위로 변화하여 기존 데이터 분석 방법으로는 분석 결과를 얻을 수 없는

매우 큰 데이터의 융합이다[4].

빅데이터의 특성인 “대용량(Volume) 데이터를 다양한 형태(Variety)와 빠른 속도(Velocity)로 가치(Value) 있는 정보를 만들어 낼 수 있는 데이터”로 정의하였다[5,6].

가치로의 변화를 통해 일부 분석 기술과 연구 방법이 필요하여 빅데이터의 특성인 볼륨, 속도 및 다양성으로 확정 지어지는 정보의 자본이라고 하였다[7].

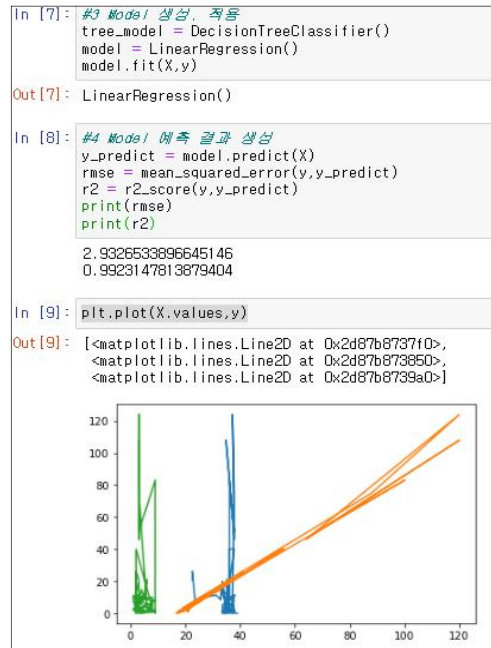


Fig. 1. Decision tree, and linear regression big data analysis results of IoT sensor

사용자와 설문조사시 인터뷰 등 정량 및 정성 연구와 같은 이전에 사용되었던 디자인 연구방법론은 제한된 표본을 바탕으로 깊게 연구를 진행하는 면에서 모든 시장과 사회의 사용자를 대변하여 상호작용과 이해의 문제를 추출하는 한계와 연구 대상이 많아짐에 따른 비교적 예산과 시간적 가치도 낮다[8].

빅데이터 분석을 활용한 정성 정량적인 최근 연구는 의사결정의 정확도를 향상하고, 비교적 근처의 미래를 예측하기 위한 융합적인 비즈니스 가치 창출에 매우 필요한 요소이다[9].

조직 구성원들 변경되고 최신 유행을 따르는 경향에서의 소셜미디어 등에 각자의 언어적 융합기술인 빅데이터 분석 활용으로 조직 변화를 이해하고 유의미한 통계적 고찰을 얻는다[10].

## 2.2 보건의복지

복지 이론은 사회 내에서 경제적, 사회적 재화의 분배에 초점을 맞춘 정치 철학의 한 분야이다. 복지는 공정성, 정의, 음식, 주거지 및 의료와 같은 기본 필수품 제공에 관한 밀접한 관련이 있습니다. 복지 이론에는 정의로운 사회를 구성하는 것과 사회 복지가 어떻게 달성되어야 하는지에 대한 고유한 관점을 가지고 있습니다. 가장 많은 사람이 가능한 한 행복하게 만들어질 때 최상의 결과를 얻을 수 있다고 주장하는 공리주의를 포함한다. 만족스러운 인생을 이어 나가는 스스로 기술과 경험에 초점을 융합한 능력 접근 방식으로 사회의 가장 혜택받지 못한 구성원의 필요를 우선시 된다. 궁극적으로 복지의 목적은 모든 사람의 기본적인 욕구가 충족되고 상품과 기회의 분배에 공정성과 정의가 있는 사회를 만드는 방법에 대한 사고의 틀이 제공된다. 최근에는 복지 기술 기업의 특수성과 복지 기술 생산과정에서의 민관협력의 실제, 정부의 구체적인 역할 등을 확인하였으며, 복지 기술 개발에서 데이터 활용에 대한 정부의 합리적 가이드라인과 규칙 제정이 필요함을 제안하였다[11].

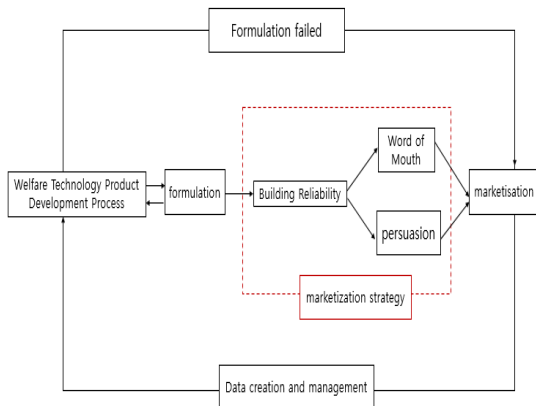


Fig. 2. Welfare technology product commercialization process

사회서비스의 혁신을 누가 어떻게 만들어가는가에 대한 이론적, 실천적 방법론으로서 '리빙랩'에 주목하였고, 리빙랩을 사용자 주도형 혁신과정이자 행위자 간 네트워크로서 이해하고 있다. 이 연구에서 복지 기술 개발의 관점에서 리빙랩의 쟁점과 유형에 관해 사례 분석한 결과, 리빙랩을 활용한 산업체, 정부, 대학교 등의 협력과 플랫폼 전략이 복지서비스의 혁신을 활성화하는 데 유용함을 확인하였다.

## 2.3 정부 정책

정책은 정부가 경제, 사회, 문화 등 분야에서 추구하는 목적을 달성하기 위해 결정하고 실행하는 계획이나 조치를 의미한다. 정책을 수립하는 데는 다양한 정책 이론이 있으며, 이는 경제학, 정치학, 행정학 등의 분야에서 발전하였다. 이러한 정책 이론은 정책 수립의 과정과 결과를 분석하고 평가하는 데 기초가 된다.

정책의 성공은 다양한 요소에 따라 평가될 수 있으며 정책의 성공은 다음과 같은 요소를 포함할 수 있습니다.

- 첫째, 목적 달성으로 정책이 설정한 목적을 달성에 성공하였는지 실패하였는지이다.
- 둘째, 실질적인 결과로써 정책이 예상한 결과와 비교하여 실질적으로 얼마나 효과적인지 한다.
- 셋째, 사회적 효과로서 정책이 사회에 미치는 영향, 예를 들어 경제적 긍정적 효과, 불평등 개선 등이 포함된다.
- 넷째, 투명성으로 정책의 제안과 실행 과정이 공정하고 투명한지를 판단한다.

마지막으로 국민의 받아들임이다. 정책이 국민에게 얼마나 수용되었는지, 예를 들어 지지율, 참여율 등으로 판단한다.

성공한 정책은 이러한 요소를 포함하고 있고, 적절한 관리와 평가를 통해 지속적으로 효과적으로 적용될 수 있다면, 각광받을 수 있을 것이다.

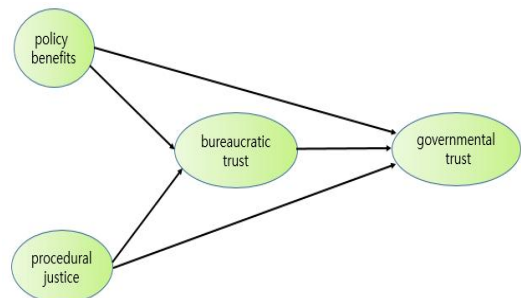


Fig. 3. Elements of Trust in Government

## 3. 분석연구설계

### 3.1 빅데이터 분석 과정

전 세계에서 모집되는 전체 데이터는 분석하는 데 어려움이 큰 형태로 가지고 있으며, 안정적이지 않은 대부분 특징이나 손실되는 값 또는 불필요하게 기록되지 않

는 데이터(noise)의 오류(error) 또는 이상치(outlier)가 확인되기에 분석에 알맞은 형태의 데이터를 최상의 형태로 변경하는 기술이 필요하다[12].

데이터 수집은 분석하고자 하는 데이터를 수집하는 단계로서 온라인 검색, 웹 스크래핑, 사용자 설문조사, 시스템 로그, 사용자 행동 분석 등 다양한 방법을 이용할 수 있다. 데이터 정제는 수집된 데이터에서 중복, 누락, 오류 등의 데이터 결함을 정제하는 단계이다. 데이터 변환은 정제된 데이터를 분석에 적합한 형태로 변환하는 단계로서 원시 데이터를 테이블 형태로 변환하거나, 텍스트 데이터를 수치 데이터로 변환할 수 있다. 데이터분석은 변환된 데이터를 분석하여 결과를 추출하는 단계로서 통계, 머신러닝, 딥러닝 등 다양한 방법을 이용할 수 있다.

### 3.2 빅데이터 분석 방법

이전의 빅데이터 분석으로는 통계적 시점에서의 확률 등으로 신뢰성 분석, 요인 분석, 상관계수를 활용한 상관분석 등 중심으로 회귀 분석용 데이터 준비과정에 크게 활용되고 있으며, 데이터가 작을 때는 딥러닝 기계학습보다 유리한 분석 기술이다[13].

4차 산업혁명시대의 빅데이터를 분석하기 위한 기계 학습분석으로서는 데이터로부터 지도학습(supervised) 비지도 학습(unsupervised) 그리고 강화 학습(reinforcement) 한 결과로 미래를 예측한다. 시뮬레이션 분석으로는 과거 혹은 미래 현상의 원인 분석 또는 가설에 대한 효과와 결과를 거시적 미시적으로 예측한다. 예를 들어 과학기술 정책 수립과 미래변화 예측 등의 방법론에 활용된다.

### 3.3 빅데이터 비정형데이터

블로그 및 소셜 네트워킹 서비스(SNS)와 같은 소셜 미디어는 고객 의견과 관련된 풍부하고 좋은 소스의 정보를 실시간 데이터로 제공한다는 측면에서 매우 효율적이며, 최근 사회적 상호작용과 정보 교환을 위한 도구로써 빠르게 제품에 대해 고객이 생생한 목소리를 전달하는 채널로써 데이터양의 폭발적 증가 추세의 측면에서 볼 때 동시대적 관점의 동향 파악에 매우 유용하다[14].

한국 언론진흥재단에서 제공하는 빅카인즈는 비정형화된 54개의 언론사의 뉴스데이터를 수집하여 뉴스 검색 분류하고 정형화된 4,919개의 빅데이터를 제공 분석한다. 데이터의 수집 기간은 2013년 02월 01일 일~2023년 02월 01일로서 10년간의 충청도 정책과 관

련된 키워드를 분석하였다.

### 3.4 빅데이터 파이선분석

파이선은 컴퓨터 그래픽 사용자들이 가장 관심을 갖는 스크립트 언어이다. 사용자가 컴파일을 하지 않고 바로 실행할 수 있으며 한 줄 단위로 실행되어서 사용자가 쉽게 결과를 확인할 수 있다는 것뿐만 아니라 파이선에서는 다양한 라이브러리를 사용 가능하다는 점 때문이다. 또한 파이선은 사용하기 쉽고 강력한 제어 능력을 제공하므로 3Ds Max, Houdini, Blender, Rhino 등 다른 3D 툴들에서도 파이선을 지원하고 있다[15]

Table 1. Big data structured in policy of Chungcheong20230201 10year CVS files

일자	언론사	기고자	제목	위치	키워드	특성추출(URL)	분석
20230131	충도일보	한성일	기거독자칼럼	화염사,고금추사,김정호옹방강,김창		http://www.jooi	
20230131	충도일보	김소희	기거세상속	대전지,대진고용,새해,2일자리,전남		http://www.jooi	
20230130	충도일보	충도일보	[사설] 철도대전,경기 증가,철도,무인점포,코			http://www.jooi	
20230129	충북일보	김기준	육전군, 아	충북남부아육전군,윤영아동학대,한		https://www.in	
20230126	충북일보	미디어전략토	키의 해, 지구,대한	토키,다산,대한민국,가		https://www.in	
20230126	충북일보	임성민	충북경찰, 충북	충북경찰,보살문조사,강		https://www.in	
20230125	충청일보	박장미	기거충북 경찰, 충북,충북	충북,경찰,충북경찰청		https://www.c	
20230119	충청일보	진재석	기거충북도의회충북,고성	충북도의회블랙리스트		https://www.c	
20230118	충청일보	충청일보	건강권 보장개성,경기건강,보장,기	본권,생활		https://www.c	
20230118	충청일보	박보성	기거예산군, 1인충남,예산	군예산군,1인예산군,공		https://www.c	
20230118	충부매일	문영호	기거野 지방의원양온전역지방,의원	(야산,공공		www.jbnews.co	
20230118	대전일보	박대항	기거예산군, 1인예산군	예산군,1인예산군,공		http://www.dae	
20230117	충청일보	충청일보	[사설] 극한만중,정주시청주시	시의회청주시의회		https://www.c	
20230117	충청일보	박보성	기거충남형 유급충남,계산,충남,유급,가	유급병,충남		https://www.c	
20230117	충부매일	황진현	기거충남도, 충충남,계산,충남도	충남유급병,거		www.jbnews.co	

## 4. 시험평가 및 결과

언론진흥재단에서 제공하는 비정형 복지정책 데이터를 CSV파일로 정형화하여 파이선 통해 Numpy, Pandas, Matplotlib등을 Impot 및 Pip install하여 키워드 코딩하고 분석하였다.

```
In [5]: kw_df = pd.DataFrame(list(kw_dict.items()), columns = ['word', 'count'])
kw_df.sort_values(by=['count'], ascending=False).head(30)

Out[5]:
```

	word	count
331	정책	9488
1124	지역	7108
342	사회	7022
1016	정부	6699
594	지원	6582
1217	교육	6044
1213	여성	5826
929	국민	4502
1240	추진	4278
1420	의원	4240
501	경제	3885

Fig. 4. Python coding keyword result

복지정책 데이터분석을 파이썬 코딩을 활용하여 빅데이터 워드 클라우드로 시각화하였다.



Fig. 5. Python coding Final result

## 5. 결론

본 연구를 기반으로 충청도 지역에서 정부의 복지정책에 대한 사회적 관심이 지역, 사회, 정부, 지원, 교육, 여성 등의 순으로 확인되어 정부에서의 복지정책에 신중히 고려를 제안한다.

빅데이터 워드 클라우드 분석 결과 가장 많은 7,108번의 언론 기사에 노출된 단어는 지역으로서 충청도 지역에서의 건강 복지정책을 중심으로 건강 관리 서비스, 의료 기관 및 의료 전문가의 접근성, 건강 교육 및 체험 프로그램 등에 정책을 검토하고, 두 번째로 많이 검색된 7,022번의 사회로서 사회 복지를 위해서 저소득 가정, 노인, 청소년, 장애인 등의 계층에 대한 지원 제공, 생활 보장 프로그램, 사회적 연대 프로그램 등에 정책에 집중 검토가 필요하다. 세 번째로 많이 검색된 6,699번의 정부로서 사회적 공제, 경제적 안정 등 다양한 분야에서 시민들의 기본 권리를 보장하는 것이 목적으로 검토해야 한다. 이외에도 6,582번의 검색된 지원, 6,044번의 검색된 교육 그리고 5,826번 검색된 여성 등도 관심이 필요하다고 판단된다. 반면 다문화, 변화, 위기, 연계, 인권, 다소 적게 검색된 것으로 분석되었다.

사회적 관심 정책을 분석한 결과 연관이 분석으로는 아동학대, 장애인, 최저임금 인상 등이 비교적 높은 연관성을 확인하였다.

## REFERENCES

- [1] Curran, C.-S., & Leker, J. (2011). Patent indicators for monitoring convergence examples from NFF and ICT. *Technological forecasting and social change*, 78(2), 256 - 273  
DOI : 10.1016/j.techfore.2010.06.021
- [2] Kim, D. Y., & Bae, J. W. (2022). Analysis of positive ESG image factors of composition using big data analysis. *Journal of the Korean Software Appraisal Society*, 18(2), 189-199.  
DOI : 10.29056/jsav.2022.12.19
- [3] Jo, J. M. (2019). Effectiveness of Normalization Pre-Processing of Big Data to the Machine Learning Performance. *The Journal of the Korea institute of electronic communication sciences*, 14(3), 547-552.  
DOI : 10.13067/JKIECS.2019.14.3.547.
- [4] Kim, D. Y., & Bae, J. W. (2022). Analysis of positive ESG image factors of composition using big data analysis. *Journal of the Korean Software Appraisal Society*, 18(2), 189-199.  
DOI : 10.29056/jsav.2022.12.19
- [5] Kim, S. H., Chang, S. H., & Lee, S. W. (2017). Consumer Trend Platform Development for Combination Analysis of Structured and Unstructured Big Data. *Journal of Digital Convergence*, 15(6), 133-143.  
DOI : 10.14400/JDC.2017.15.6.133
- [6] Wang, J., Ma, Y., Zhang, L., Gao, R. X., & Wu, D. (2018). Deep learning for smart manufacturing Methods and applications. *Journal of Manufacturing System*, 48(2018), 144-156
- [7] De, M. A., Greco, M., & Grimaldi, M. (2016). A formal definition of Big Data based on its essential features. *Library Review*, 65(3), 122-135
- [8] Lee, H. J. (2021). Data-Driven Design Methodology based on Data Science Paradigm Focused on Design Research Case Study of Fine Dust Information App Service. *The Journal of the Korea Contents Association*, 21(10), 103-114.  
DOI : 10.5392/JKCA.2021.21.10.103
- [9] Park, S. H. (2018). Design Thinking Methodology for Social Innovation using Big Data and Qualitative Research. *Asia-Pacific journal of business and venturing*, 13(4), 169-181.  
DOI : 110.16972/apjvbe.13.4.201808.169
- [10] Jung, S. M. (2012). A Study on the Classification system and fundamental notions of the Peripheral Devices and Accessories in Product Design for



Smartphone. *Journal of Korea design knowledge*, 24(1), 20-29.

DOI : 10.17246/jkdk.2012..24.003

- [11] Choi, J. H., Lim, J. W., & Kim, S. W. (2022). A study on the commercialization process of welfare technology in the field of health and welfare. *Journal of Korean Social Welfare Administration*, 24(3), 7-67.  
DOI : 10.22944/KSWA.2022.24.3.002
- [12] Kim, D. Y., & Bae, J. W. (2022). Analysis of positive ESG image factors of composition using big data analysis. *Journal of the Korean Software Appraisal Society*, 18(2), 189-199.  
DOI : 10.29056/jsav.2022.12.19
- [13] Kim, D. Y., & Bae, J. W. (2022). Analysis of positive ESG image factors of composition using big data analysis. *Journal of the Korean Software Appraisal Society*, 18(2), 189-199.  
DOI : 10.29056/jsav.2022.12.19
- [14] Kim, D. Y., & Bae, J. W. (2022). Analysis of positive ESG image factors of composition using big data analysis. *Journal of the Korean Software Appraisal Society*, 18(2), 189-199.  
DOI : 10.29056/jsav.2022.12.19
- [15] Mechtley, A., & Trowbridge, R. (2012). Maya Python for games and film a complete reference for the Maya Python and the Maya Python API. Morgan Kaufmann. *Amsterdam Boston*, 24(1), 20-88.  
DOI : 10.9708/jksci.2015.20.4.049

김 대 유(Dae-Yu Kim)

[정회원]



- 2012년 2월 : 한국산업기술대학교 IT융합(공학사)
- 2015년 2월 : 서강대학교 기술경영 (기술경영석사)
- 2020년 8월 ~ 현재 : 한국기술교육대 기술경영(박사 수료)

- 2021년 3월 ~ 현재 : 남서울대학교 빅데이터산업보안 조 교수
- 관심분야 : 통계, 빅데이터, 기술경영, 창업가정신
- E-Mail : dy.kim@nsu.ac.kr

나 원 식(Won-Shik Na)

[종신회원]



- 2005년 8월 : 경희대학교 컴퓨터공학과 (공학박사)
- 2020년 9월 ~ 현재 : 소프트웨어(SW) 미래채움 충남사업 단장
- 2022년 3월 ~ 현재 : 남서울대학교 원격평생교육원 원장

- 2006년 3월 ~ 현재 : 남서울대학교 컴퓨터소프트웨어학과 교수
- 관심분야 : 네트워크 보안, 정보보호, 무선 LAN, 의료정보, 전자제어 등
- E-Mail : winner@nsu.ac.kr