

Text Based Explainable AI for Monitoring National Innovations

Jung Sun Lim*[†] · Seoung Hun Bae**

*KISTI, **LXSIRI

텍스트 기반 Explainable AI를 적용한 국가연구개발혁신 모니터링

임정선*[†] · 배성훈**

*한국과학기술정보연구원, ** 한국국토정보공사 공간정보연구원

Explainable AI (XAI) is an approach that leverages artificial intelligence to support human decision-making. Recently, governments of several countries including Korea are attempting objective evidence-based analyses of R&D investments with returns by analyzing quantitative data. Over the past decade, governments have invested in relevant researches, allowing government officials to gain insights to help them evaluate past performances and discuss future policy directions. Compared to the size that has not been used yet, the utilization of the text information (accumulated in national DBs) so far is low level. The current study utilizes a text mining strategy for monitoring innovations along with a case study of smart-farms in the Honam region.

Keywords : XAI, Explainable AI, Innovation, Monitoring

1. 서 론

미국과 유럽연합을 포함한 세계 주요국 정부들은 객관적인 데이터 및 분석결과에 기반하여 정책을 기획하고, 이행하며, 평가하기 위한 수단을 필요로 한다. 각국 정부들은 이를 위한 증거기반(Evidence Based) 성과측량체계(Innovation Measurement Metric)를 연구하고, 유관 플랫폼을 구축할 것을 공공과 연구계에 요청하여 왔다. 예로서 2007~2008년 경제 위기 극복을 위해 제정된 미국의 ARRA(American Recovery and Reinvestment Act) of 2009 법에는, 신규 연구개발투자에 의한 고용성과를 증거기반으로 측정할 것을 요구하고 있다. 본 법에 대응하여, 백악관 과학기술정책실(Office of Science and Technology Policy)을 중심으로 STAR METRICS (Science and Technology for America's Reinvestment: Measuring the Effect of Research on Innovation, Competitiveness and Science)라는 이름의 연구개발혁신 측정 연구를

수행하고[22] 또한 유관 플랫폼을 구축[23] 하였다. 유럽연합 역시 미국연방과 유사한 시기에 발생한 경제위기 극복을 위해 연방차원의 연구개발투자 전략들을 발표한다. 유럽연합의 통합 연구개발체계 Horizon 2020(Framework Programme 8)은 연방이 집중관리하는 산업적 전략기술군들 KETs(Key Enabling Technologies)을 육성하는 Industrial Leadership 프로그램을 신설하였다. 또한 KETs 들이 창출하는 혁신성과들을 증거기반 플랫폼을 통해 분석하고 모니터링하기 위한 목적의 KETs Observatory 체계를 연구하고[9] 또한 유관 데이터분석 플랫폼을 구축하였다[8]. Horizon 2020을 계승하는 연방의 차기 연구개발체계인 Horizon Europe(FP 9) 하에서는 ATI 플랫폼(KETs Observatory를 개량한 체계)을 통해 유럽연합이 산업적으로 주목하는 전략 기술군들(본 연구의 사례분석 예인 스마트팜 포함)의 혁신 성과들을 증거기반으로 모니터링하여 관련 정보들을 공공 및 민간에 제공하여 왔다[11]. 요약하면 미국 및 유럽연합은 국가가 육성중인 산업화 전략기술군들이 경제 및 사회에 창출하는 혁신(Innovation) 성과들을 객관적으로 측정(Measurement)하기 위한 증거기반 방법론을 다년간 연구하고

Received 8 September 2022; Finally Revised 7 November 2022;
Accepted 9 November 2022

[†] Corresponding Author : jsunnylim@kisti.re.kr

또한 유관 데이터 플랫폼을 구축하여 왔다. 참고로 OECD가 다년간 연구 중인 혁신 및 혁신측정 방법론 Oslo Manual[31]에 의하면, 혁신은 “기존 대비 개선되거나 새로운 제품 및 프로세스를 창출하는 활동”으로 정의된다. 앞서 언급한 미국 및 유럽연합의 혁신측량 플랫폼들, 그리고 이하 소개할 한국 R&D PIE에 활용되는 혁신의 의미는 모두 유사하다.

앞선 미국과 유럽연합 사례와 같이, 대한민국 정부 역시 산업적으로 가치가 높은 전략기술군들을 설정하고 또한 이들의 경제적 사회적 혁신 성과들을 모니터링하는 증거 기반 플랫폼을 구축할 것을 공공 및 연구계에 요청하여 왔다. 그 결과물로, 국가연구개발투자 및 성과DB(NTIS, ntis.go.kr[30])로 부터 대한민국정부가 주목하는 10대 전략 분야를 도출한 R&D PIE(Platform for Investment and Evaluation)[10, 26] 플랫폼을 구축하였다. 2018년 이후부터 R&D PIE는 정책당국자들의 국가연구개발예산 배분과정을 지원하기 위한 데이터들을 분석하기 위한 플랫폼으로 활용되어 왔다. 참고로 R&D PIE는 정부가 그 산업적 가치를 인정한 10대 전략기술군들의 정보를 담고 있기에, 그 유관 분석결과들은 공공분야에 뿐만 아니라 민간산업 분야에서도 활용 가치가 높다. R&D PIE는 경제장관회의[12]를 통해 국가의 연국개발투자 배분과정에 활용될 것임이 발표되거나, OECD의 세계 공공부문 혁신사례[32]로서 선정되기도 하였다.

R&D PIE는 정책당국자들의 활동을 지원하는 용도 뿐만 아니라, 유관 기능 고도화를 위한 탐색연구들도 수행되어 왔다. R&D PIE DB를 구축하기 위해 유관 정보들을 NTIS 원시DB(R&D PIE DB 포함)로 부터 재조합 하고 마이닝 하는 자동화 플랫폼을 구축하고, 또한 이를 기반하여 국가연구개발투자가 창출한 혁신을 측량하는 연구들을[24-27] 수행 하였다. 본 플랫폼을 통해 부처 혹은 지역간 혁신 특성을 구분하고[24-26], 고용 및 경제유발 효과를 추정하는[27] 연구들을 수행하였다. 이들의 연장선에서, 본 연구논문은 그동안 NTIS DB내에 내장되어 있으나 그 활용이 미비 하였던 텍스트 정보를 마이닝하는 분석체계를 탐색연구 하였다.

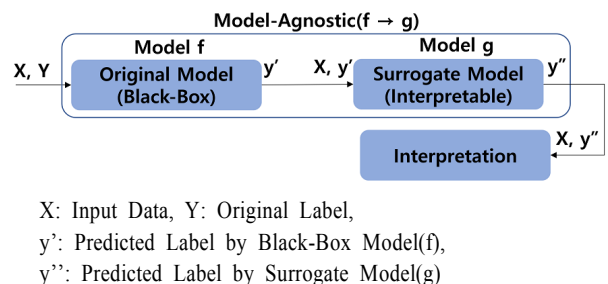
본 논문은 한정된 지면의 한계를 고려하여, R&D PIE의 10대 전략분야 중 스마트팜[12]을 모델분석 대상으로 삼는다. 본 스마트팜 DB내에 포함되어 있었으나, 그 활용이 매우 제한적이었던 비정형 텍스트에 기계학습(Machine Learning, 이하 ML)과 AI 해석(Explainable Artificial Intelligence, 이하 XAI) 기법을 적용한 분석체계를 탐색연구 하였다. 앞서 소개한 유럽연합[9-11, 19] 및 미국연방[22-23] 등의 국가혁신 모니터링 연구사례들 중 텍스트 비정형 데이터의 활용은 여전히 초기단계이며, 최근[11]에 들어 키워드 네트워크 분석연구가 진행 중이다. KETs Observatory와 같은 국가DB 기반의 모니터링 시스템에서 파생된

논문사례들[8, 13, 15, 29, 33]은, 주로 경영과학 및 다학제 연구를 통해 정부의 정책방향 수립에 기여하고 또한 첨단 산업기술의 경제 사회적 기여를 분석하는 경우들 이었다. 그러나 본 연구논문과 같이 텍스트 기반의 기계학습 및 해석 기술이 적용된 연구된 사례는 현재까지 파악되지 않고 있다. 본 연구는 정책현장의 데이터인 R&D PIE 스마트팜 분야를 사례로, 텍스트 기반의 기계학습 해석(XAI) 기법을 적용한 정보획득 방법론을 탐색하였다. 스마트팜(첨단 농업)에 의한 생산품들과 핵심어(유관 국가과학기술표준 분류, 핵심기술 등) 정보를 기계학습의 도움을 통해 요약 및 해석하는 사례 연구를 수행하였다.

2. 선행 연구

기계학습의 성능이 향상될수록 모델은 Black-Box로 표현될 만큼 해석이 어려워지며, 기계학습이 결과를 도출하게 된 근거를 역추적하는 기술들을 필요로 하게 되었다. 기계학습 해석 기법들은 미 국방성 DARPA 프로젝트 XAI (Explainable AI)[6]를 통해 세계적인 주목을 받았다. 기계학습 해석 관련 선행연구들은 악성코드 예측[21], 스마트팩토리[6], 부도율예측[1] 등이 파악된다. 활용되는 기계학습 모형들로 RF(Random Forest), XGB(Extreme Gradient Boosted), SVM(Support Vector Machine), Neural Network (LSTM, BERT) 등이 파악되며 AI해석 기법에는 LIME, SHAP, LRP, GLM, RF 등이 파악된다[1-2, 4-7, 14, 16-18, 20-21, 28].

정책현장의 지원업무는 요청받는 데이터 분석의 규모를 사전에 예측하기 어려우며, 또한 요청받는 대응시간이 짧은 경우가 빈번하다. 앞서 소개한 텍스트기반 선행연구들과 같은 높은 수준의 연산자원과 긴 처리 시간을 필요로 하는 경우와는 별개로, 현실적인 연산 시간 내에서 유연한 적용과 활용이 가능한 기계학습 및 해석 체계를 탐색연구 하였다. Explainable AI의 한 종류인 Model-Agnostic[3, 28] 기법을 활용하여 Black-Box 모형에 R&D PIE(NTIS DB 기반) 텍스트



<Figure 1> Work-flow Image of Model-Agnostic Approach [3, 28]

정보를 학습시킨다. 그리고 Black-Box 모형의 예측결과를 가장 잘 모사(Emulation)하는 해석 가능한(Explainable, Interpretable) 대리모형(Surrogate)을 통해 Black-Box 모형의 판단 근거를 해석한다. 참고로 Model-Agnostic[3, 28]은 기계학습 모형들 간 종류에 구애받지 않고 상호 연계될 수 있다는 철학을 가진 분석 방법론이다.

3. 연구방법

3.1 분석 데이터

분석의 모델 데이터는 미래창조과학부 혁신본부의 정책담당자들 및 이들을 지원하는 기술전문그룹들의 협력을 통해 도출된 R&D PIE[12], 그중 스마트팜 분과의 연구개발투자 과제리스트를 기초자료로 사용하였다. 또한 대한민국정부의 연구개발투자 및 성과정보를 망라한 NTIS 원시DB 중, 국가의 조사분석평가가 완료된 가장 최신정보(연구개발투자 2010~2020년(613,205건), 사업화성과 2016~2020년(148,900건))를 ntis.go.kr[30]를 통해 다운로드 받아 사용하였다.

본 연구에서 활용하는 비정형 텍스트들은 연구개발투자 DB(국문과제명, 한글키워드, 지역분류), 그리고 사업화 DB(사업화명, 사업화내용)를 조합(Concatenate) 후 Mecab을 통해 자연어를 전처리하였다. 이하 분석에서 사용되는 생산품 리스트는 사업화 DB중 “제품명” 컬럼정보를 참고하여 사업화명/상품명 필드 중 유관한 키워드들을 분석자가 수작업으로 선별하였다. 또한 본 연구에서는 스마트팜 사업화 성과(전국기준 총 1199건) 중, 지면의 제약상 호남권(전라북도, 전라남도, 광주광역시; 328건)을 예로 기계학습과 해석,

그리고 네트워크분석을 수행하였다. 기계학습을 위한 분류 기준(Labeling)으로 국가의 연구개발표준분류 6T(IT, BT, NT, ST, ET, CT, etc(기타 미분류))를 사용하였다. 기계학습에 활용되는 텍스트는 총 2,986개의 Bag-of-Word 컬럼, 그리고 7개의 분류컬럼(6T 및 etc)으로 변환하여 사용하였다.

3.2 기계학습 및 해석 방법론

본 연구는 <Figure 1>과 <Figure 2>에 요약되었듯이 데이터 전처리, Black-Box 기계학습, Black-Box 모형을 모사하기 위한 대리학습, 그리고 대리학습 모형의 결과를 해석하는 해석 과정으로 워크플로우를 구성하였다.

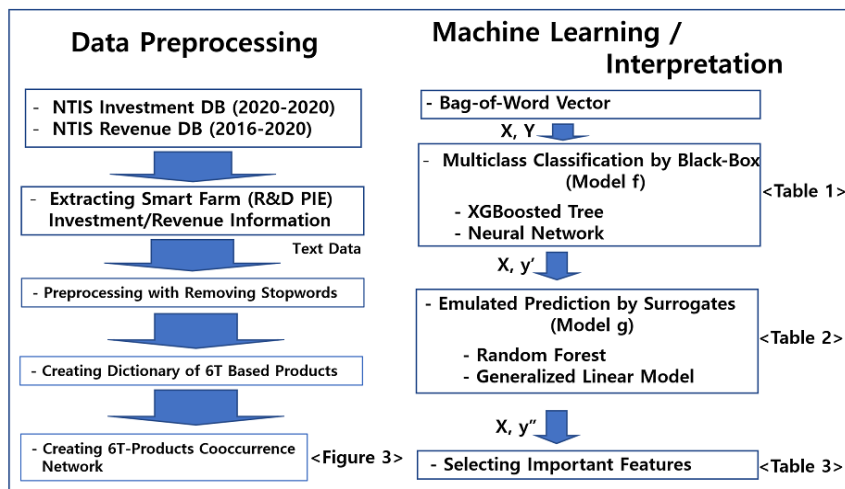
본 연구는 학습속도가 비교적 빠르며 범용적인 Black-Box 모형들간 성능을 Accuracy 기준으로 상호 비교 후 선택하였다. Black-Box 모형의 예측결과는 해석 가능한 모형들에 전달되며, 그 중간분석 결과를 6T 분류기준의 오차행렬 (Confusion Matrix)로 <Table 1>~<Table 2>에 정리하였다. 오차행렬은 분류별 TP(True Positive), TN(True Negative), FP(False Positive), FN(False Negative) 지표들을 통해 재현율 (Recall, 실제 정답 중 모델이 정답을 예측한 비율), 정밀도 (Precision, 모델이 예측한 정답 중 실제 정답 비율), F1(Recall 과 Precision의 조화평균), 정확도(Accuracy)를 측정한다.

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1 - measure = \frac{2P \times R}{P + R}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$



<Figure 2> Overall Research Frame

<Table 1>에 요약되었듯이 Black-Box 모형중 가장 우수한 성능을 보인 Keras 기반의 Multilayer Perceptron 모델이 이하 분석에서 사용되었으며 Input Layer, Reshape Layer, Dense Layer(ReLU activation), Dense Layer 및 Adam Optimizer로 구성되어 있다. 이하 분석에서 직접 활용되지는 않았으나 <Table 2>~<Table 3>에 언급된 XGBoost는 빠른 처리속도와 우수한 성능을 보이는 대표적인 나무(CART, Classification and Regression Tree) 기반 알고리즘으로, 다수의 약한 예측 모형들을 직렬(Boosting) 형태로 결합하여 보다 오차를 줄이는 예측모형을 구축하는 앙상블 기법이다[6, 21]. Random Forest는 또 다른 대표적인 나무기반 알고리즘으로, 다수의 약한 예측모형들을 상호 독립적인 병렬(Bagging) 형태로 결합하여 보다 강한 예측모형을 구축하는 앙상블 기법이다 [4, 14, 21]. Random Forest는 또한 Feature(입력변수 혹은 독립변수)의 중요도를 분석하는 방법[4, 14]으로도 활용된다. <Table 3>와 같이 GLM(Generalized Linear Model) 모형은 Keras MLP 모형의 예측결과를 가장 유사하게 모사하였으며, GLM을 통해 Feature(입력변수 혹은 독립변수)의 중요도를 분석하는데 활용하였다. GLM[17-18]을 AI해석을 위한 대리모형으로 활용하는 사례연구를[18] 참고였다. H₂O사가 공개한 GLM분석 패키지[17]를 사용하였으며 6T(국가과학기술분류)별 “Multinomial” 모드를 통해 다중분류(Multiclass Classification)를 수행한다. GLM을 통해 도출된 Feature별 계수(Coefficient) 값들을 중요도로 파악하여, 상위 20개 단어를 선택하였다.

3.3 동시출현 네트워크

스마트팜 연구개발투자에 의해 매출을 발생시킨 과제들을 기준으로 NTIS 연구개발투자DB 내의 자연어(국문 과제명, 한글키워드, 6T분류), 그리고 성과DB 내의 자연어(사업화명, 사업화설명)를 병합하여 동시출현 네트워크를 생성한다. 생성된 네트워크에, 앞서 분석자가 수작업으로 작성한 스마트팜 생산품 리스트를 사용해 필터링하여 6T-생산품 동시출현 네트워크를 작성한다. 가시성의 한계를 고려하여, 본 네트워크중 상위 30개 Node를 선별하여 <Figure 3>에 이미지화 하였다. 또한 위 언급한 생산품 리스트는 인공지능 대리모형을 해석하는 과정에서 생산품과 관련한 핵심 키워드들을 필터링 하는데도 활용하였다.

4. 분석 및 결과

6T로 분류된 스마트팜 사업화성과 전체 데이터를 8:2 (학습:평가) 비율로 다양한 모델에 대해 학습 및 평가하여, 가장 우수한 결과(Accuracy)를 도출한 XGBoost 및 Keras MLP의 오차행렬을 <Table 1>에 요약하였다. XGBoost 대비 Keras MLP의 예측결과가 ET 분류를 기준으로 보다 우수한 성능을 보였다. Keras MLP에 의한 학습모델(f)을 이하 분석에서 사용하였다.

본 분석에서는 스마트팜 성과중 호남권(전라남도, 전라

<Table 1> 6T Based Classification Results (y') by Black-Box Models

	XGBoost				Keras Multilayer Perceptron (Keras MLP)			
	Recall	Precision	F-1	Accuracy	Recall	Precision	F-1	Accuracy
BT	0.98	0.90	0.94	0.92	1.00	0.97	0.99	0.98
CT	1.00	1.00	1.00		1.00	1.00	1.00	
ET	0.52	1.00	0.69		0.90	1.00	0.95	
etc	0.95	0.93	0.94		1.00	0.98	0.99	
IT	0.90	0.91	0.91		0.97	0.90	0.98	
NT	1.00	1.00	1.00		1.00	1.00	1.00	
ST	1.00	1.00	1.00		1.00	1.00	1.00	

<Table 2> Emulation Results (y'') by Interpretable Models

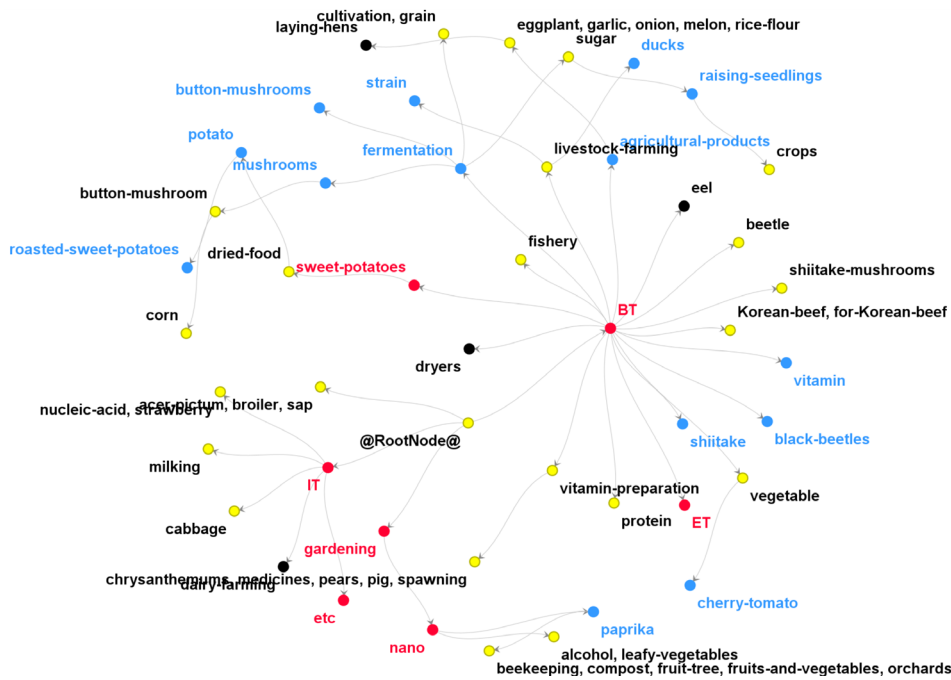
	Random Forest (RF)				Generalized Linear Model (GLM)			
	Recall	Precision	F-1	Accuracy	Recall	Precision	F-1	Accuracy
BT	1.00	0.54	0.71	0.57	1.00	1.00	1.00	1.00
CT	1.00	1.00	1.00		1.00	1.00	1.00	
ET	-	-	-		1.00	1.00	1.00	
etc	0.06	1.00	0.12		1.00	1.00	1.00	
IT	-	-	-		1.00	1.00	1.00	
NT	0.4	1.00	0.57		1.00	1.00	1.00	
ST	0.78	1.00	0.88		1.00	1.00	1.00	

북도, 광주광역시)을 기계학습해석의 사례로, 그 과정을 <Figure 2>에 요약하였다. <Table 1>은 Keras 기반 Black-Box 모델의 예측결과(y')가 원본데이터의 라벨(Y)을 얼마나 유사하게 재현하였는가를 분석한 오차행렬 결과이다. <Table 2>는 대리분석(Surrogate) 모델의 예측결과(y'')가 Black-Box 모델의 예측결과(y')를 얼마나 잘 재현하였는가를 분석한 오차행렬 결과이다. 해석부의 GLM 대리분석 모형에 의한 오차행렬 결과(<Table 2>)가 가장 우수하였으며, 앞서 언급하였듯이 H₂O의 Multinomial GLM 함수에 의하여 6T분류별 Feature(입력변수 혹은 독립변수)의 계수(Coefficient)를 중요도로 사용하였다.

앞서 작성된 호남권의 6T → 생산품 동시출현 네트워크 중, 생산품과 링크된 키워드들만을 분리 후 추출하였다. 본 키워드들은 다시 GLM 대리분석을 통해 도출된 Feature 계수 테이블 결과를 필터링 하는데 적용하여, 6T별 상위 20개의 키워드들을 필터링 후 <Table 3>에 정리하였다.

NT(나노) 분야의 네트워크분석 결과 원예, 파프리카,

잎채소 등의 생산품이 모니터링 되었다. NT의 기계학습 해석 결과를 통해 광원(엘이디)이 중요 키워드로 파악되었는데, 엘이디는 스마트팜의 시설원예, (수경 및 인공) 재배 등에 필요한 인공광원이다. Etc(기타 미분류)의 네트워크분석 중 양계사업(산란 및 육계)이 모니터링 되었으며, 기계학습해석 과정 중 친환경 키워드들이(악취, 제어기) 파악되었다. 대부분의 생산품은 IT 및 BT에 기반한 노지 농업(마늘, 버섯, 고구마, 딸기, 과수원 등), 낙농(한우, 착유 등), 양계, 양식 사례들이었다. 기계학습해석을 통해 BT분류 기준으로 재배기, 사물인터넷, 신선, 건조, 새싹 등의 키워드가 중요한 것으로 파악되었다. 또한 기계학습해석을 통해 IT분류 기준으로 관수, 이력, 양액, 지능화, 클라우드, 관측 등의 키워드가 중요한 것으로 파악되었다. BT 및 IT 기술군들이 스마트팜 기반 생산활동의 초기(파종, 새싹 등), 중간(관수, 이력, 지능화, 예측 등), 최종(작황, 재배기 등) 과정에 이르는 전주기 스마트화에 핵심적으로 활용되는 것으로 파악된다.



<Figure 3> Co-Occurrence Network Related with 6T-Based Products

<Table 3> Selected 6T-Based Keywords by AI Interpretation

BT	Cultivator, Sowing, Spray, High Speed, Synchronous, Fresh, Standard, Objects, Internet, Progress, Machine, Sprout, Breathable, Shinma, Consumption, Drying, Bulk, Processing
IT	Irrigation, Convenience, Weight, Precision, Cloud, Intelligence, Wireless, History, Lora, Nutrient-Solution, Sorting, Multi-Item, Energy-Source, Ultrasound, Observation, Large-Size, Cropping, Prediction, Molecule, Label
NT(Nano)	LED, Lighting, Livestock, Plant, Sterilization, Prevention, Barn, Function, Cultivation, Harmful, Sensor, Agriculture, Ammonia, Applicable, Carbon, Vulnerable, Indicator
etc	Controller, Odor, Small, Semi-Automatic, Self-Reliant, Bio, Open, Underground, Working-Machine, Harvester, General-Purpose, Chassis, Transmission, Planter, Light-Emission, Resolution, Domestic, Indicator, Equipment

5. 결론

국내외 주요 정부들은 국가의 연구개발투자 및 성과 DB를 분석하기 위한 플랫폼을 연구하고 또한 유관 시스템을 구축할 것을 공공 및 연구계에 요청하여 왔다. 정책당국자들은 연구개발된 결과물 및 유관 시스템 들을 정책현장에서 활용하고자 한다. 비정형 텍스트 정보의 중요성이 널리 알려져 있음에도 불구하고, 국내외 정책현장에서의 활용은 여전히 제한적이다. 본 연구는 유관 현장에서 텍스트 기반의 기계학습 및 해석 방법론과 같은 진보된 정보마이닝 기법이 활용될 수 있는 기반 마련을 위해 수행되었다. 대한민국정부의 연구개발투자 및 성과 DB(NTIS DB), 이를 기반해 구축된 R&D PIE 중 스마트팜 분야를 사례로 일련의 분석체계를 모델연구 하였다. 호남권에서 유발된 다양한 생산품 정보들을 동시출현 네트워크를 통해 요약하고, 생산품과 깊이 연관된 키워드들을 기계학습 해석 방법론으로 추론하는 분석체계를 시연하였다. 본 연구는 현장의 제한된 시간 및 자원 내에서 분석자가 시스템의 보조하에 다각도의 정보를 마이닝 하고, 또한 최종판단에 집중할 수 있는 환경 마련에 기여하고자 수행되었다.

Acknowledgement

This study was supported by research funds of KISTI(Korea Institute of Science and Technology Information) K-22-L05-C02-S07 with K-22-L03-C04-S01.

References

- [1] Ahn, H. and Lim, D.H., A Study on Fraud Detection in the C2C Used Trade Market Using Doc2vec, *Journal of The Korea Society of Computer and Information*, 2022, Vol. 27, No. 3, pp. 173-182.
- [2] Angelov, P.P., Soares, E.A., Jiang, R., Arnold, N.I., and Atkinson, P.M., Explainable Artificial Intelligence: An Analytical Review, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2021, Vol. 11, No. 5, pp. 1-13.
- [3] Arrieta, A. B., Diaz-Rodriguez, N., Ser, J. D., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molia, D., Benjamins, R., Chatila, R., and Herrera, F., Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges Toward Responsible AI, *Information Fusion*, 2020, Vol. 58, pp. 82-115.
- [4] Cao, D.S., Lianga, Y.Z., Xub, Q.S., Zhanga, L.X., Huc, Q.N. and Lia, H.D., Feature importance sampling-based adaptive random forest as a useful tool to screen underlying lead compounds, *J. Chemometrics*, 2011, Vol. 25, pp. 201-207.
- [5] Chen, B.C., Kang, K.M., and Kim, J.K., Deep Learning-based Text Summarization Model for Explainable Personalized Movie Recommendation Service, *Journal of Information Technology Services*, 2022, Vol. 21, No. 2, pp. 109-126.
- [6] Cheon, K.M. and Yang, J., Explainable AI Application for Machine Predictive Maintenance, *Journal of Korean Society of Industrial and Systems Engineering*, 2021, Vol. 44, No. 4, pp. 227-233.
- [7] DARPA., Explainable Artificial Intelligence(XAI), 2022, [Online]. Available: <https://www.darpa.mil/program/explainable-artificial-intelligence>.
- [8] Duskova, M. and Souckova, M., Company and State Performance in Key Enabling Technologies-Case Study of Nanotechnology Industry, *International Journal of Innovation and Technology Management*, 2020, Vol. 17, No. 1, 2050002.
- [9] EC(European Commission), Feasibility study for and EU monitoring mechanism on key enabling technologies, Belgium: European Commission, 2012.
- [10] EC, *KETs Obsevatory Phase ii Methodology Report*, Belgium: European Commission, 2017.
- [11] EC, *Advanced Technologies for Industry*, 2022, [Online]. Available: <https://ati.ec.europa.eu/>.
- [12] Economic Ministers' Meeting 18-3, Government R&D investment innovation plan to support innovative growth, Joint ministry of Korea Government, 2018.
- [13] Evangelista, R., Meliciani, V., and Vezzani A., Specialisation in key enabling technologies and regional growth in Europe, *Economics of Innovation and New Technology*, 2018, Vol. 27, No. 3, pp. 273-289.
- [14] Fabris, F., Doherty, A., Palmer, D., Pedro de Magalhaes, J., and Freitas, A., A New Approach for Interpreting Random Forest Models and Its Application to the Biology of Ageing, *Bioinformatics*, 2018, Vol. 34, No. 14, pp. 2449-2456.
- [15] Gwarda-Gruszczynska, E. and Dobrowolska, B., KETs Diffusion and its Influence on Companies' Innovativeness and Competitiveness, *Economics and Organization of the Enterprise*, 2017, Vol. 6, pp. 47-57.
- [16] Ha, T., Coh, B.Y., Lee, M., Yun, B., and Chun, H.W., An Exploratory Approach to Discovering Salary-Related

- Wording in Job Postings in Korea, *Journal of Information Science Theory and Practice*, 2022, Vol. 10, pp. 86-95.
- [17] Hall, P., Gil, N., Kurka, M., and Phan, W., Machine Learning Interpretability with H₂O Driverless AI, 2022, [Online], Available: <https://docs.h2o.ai/driverless-ai/latest-stable/docs/booklets/MLIBooklet.pdf>.
- [18] Henckaerts, R., Antonio, K., and Cote, M.P., When Stakes are High: Balancing Accuracy and Transparency with Model-Agnostic Interpretable Data-driven suRRogates, *Expert Systems with Applications*, 2022, Vol. 202, pp. 1-13.
- [19] Horvat, M., The New Framework for EU Research and Innovation, *Science*, 2011, Vol. 334, pp. 1066-1068.
- [20] Hwang, H. and Lee, Y., Semi-Supervised Learning based on Auto-generated Lexicon using XAI in Sentiment Analysis, *Proceedings of Recent Advances in Natural Language Processing*, 2021, pp. 593-600.
- [21] Kim, D.Y., Jeong, A.Y., and Lee, T.J., Analysis of Malware Group Classification with eXplainable Artificial Intelligence, *Journal of The Korea Institute of Information Security & Cryptology*, 2021, Vol. 31, No. 4, pp. 559-571.
- [22] Lane, J., Let's Make Science Metrics More Scientific, *Nature*, 2010, Vol. 464, pp. 488-489.
- [23] Lane, J. and Bertuzzi, S., The Star Metrics Project: Current and Future Uses for S&E Workforce Data, 2010, [Online], Available: <https://www.nsf.gov/sbe/sosp/workforce/lane.pdf>.
- [24] Lim, J.S., Shin, K.M., Yoon, J.S., and Bae, S.H., Study of US/EU National Innovation Policies Based on Nanotechnology Development, and Implications for Korea, *Journal of Information Science Theory and Practice*, 2015, Vol. 3, No. 1, pp. 50-65.
- [25] Lim, J.S. and Kim, S.G., Quantifying Innovation Characteristics of National R&D Investment through Cross-Cutting Data Analysis: Case Study on NTIS Nanotechnology, *Journal of The Korean Operations Research and Management Science Society*, 2020, Vol. 45, No. 2, pp. 13-23.
- [26] Lim, J.S., Kim, S.G., Bae, S.H., Kim, K.H., and Won, D.K., Research on Regional Spatial Information Analysis Platform about NTIS Raw Data, *Journal of Cadastre & Land InformatiX*, 2020, Vol. 50, No. 2, pp. 21-35.
- [27] Lim, J.S., Park, J.H., and Kim, S.G., Study on Measurement of National R&D Investment Spillover: Linking Multilateral Big Data, *Korea CEO Academy*, 2021, Vol. 24, No. 1, pp. 311-330.
- [28] Messalas, A., Kanellopoulos, Y., and Makris, C., Model-Agnostic Interpretability with Shapley Values, *10th International Conference on IISA*, 2019, pp. 1-7.
- [29] Mieszkowski, K., The Methods and Rationale for Selecting Advanced Manufacturing Technologies as a Smart Specialisation Priority in Less-innovative Countries: The Case of Poland, *Problemy Eksploatacji*, 2016, Vol. 2, pp. 71-82.
- [30] NTIS(National Science & Technology Information Service), 2022, [Online]. Available: <https://www.ntis.go.kr>.
- [31] OECD, Oslo Manual 2018, Guidelines for Collecting, Reporting and Using Data on Innovation, OECD Publications, 2018.
- [32] OECD, R&D Platform for Investment and Evaluation ("R&D PIE"), 2022, [Online], Available: <https://oecd-opsi.org/innovations/rd-platform-for-investment-and-evaluation-rd-pie>.
- [33] Salikhova, O.B., Industry Modernization Based on the Principles of Smart Specialization. Part II. Establishing the Tools for Monitoring and System Analysis of the Impact of Key Technologies on Industry Modernization, *Statistics of Ukraine*, 2020, Vol.88, No.1, pp. 84-90.

ORCID

Jung Sun Lim | <https://orcid.org/0000-0002-2737-6118>

Seoung Hun Bae | <https://orcid.org/0000-0002-0819-4386>