Towards administrative innovation in the public sector: A policy informatics approach

Graduate School of Management of Technology
Sogang University

Changyong Lee, Ph.D.
Email: changyong@sogang.ac.kr

Contents

I. Introduction

1. About me
2. Policy informatics
3. AI and machine learning

II. Expert-machine collaboration for decision making

1. Research background and problem
2. Data and methodology
3. Empirical analysis and results
4. Potential applications

III. Mismatch between the supply and demand

1. Research background and problem
2. Data and methodology
3. Empirical analysis and results
4. Potential applications

IV. Opinion mining, sentiment analysis, and simulation

1. Opinion mining for agenda setting
2. Sentiment analysis for agenda setting and policy evaluation
3. Simulation for policy evaluation
4. Chatbot for policy implementation

V. Conclusion

1. Challenges and paths to the next stage

I. Introduction

1. About me

Education

- PhD in Industrial Engineering
- BS in Computer Science (Minor in Industrial Engineering)

Work experience

- Ulsan National Institute of Science and Technology
- Korea Institute of Science and Technology Information
- Centre for Technology Management, University of Cambridge

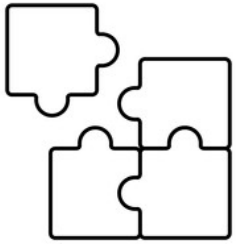
Research interests

- **Policy informatics and data-driven policy evaluation** (e.g., forecasting, evaluation, and planning in the public sector)
- Industrial policy (e.g., sustainability of small and medium enterprises)
- Applied machine and deep learning (e.g., prognostics and health management of complex systems and high-frequency trading in market-making contexts)

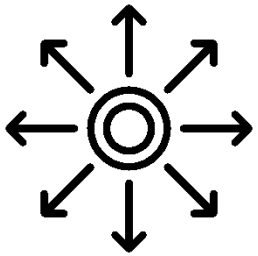
1. About me

- **Major research topics**

[Identification]
Customer and market needs and emerging technologies

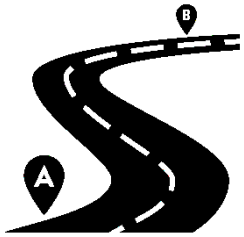


[Association/ Matchmaking]
Disparate databases and supply and demand



[Alternatives/ Recommendation]
Products, services, technologies, businesses, and scenarios

[Evaluation/ Decision making]
Attractiveness, risk, uncertainty, and satisfaction



[Planning]
Technology and business roadmaps

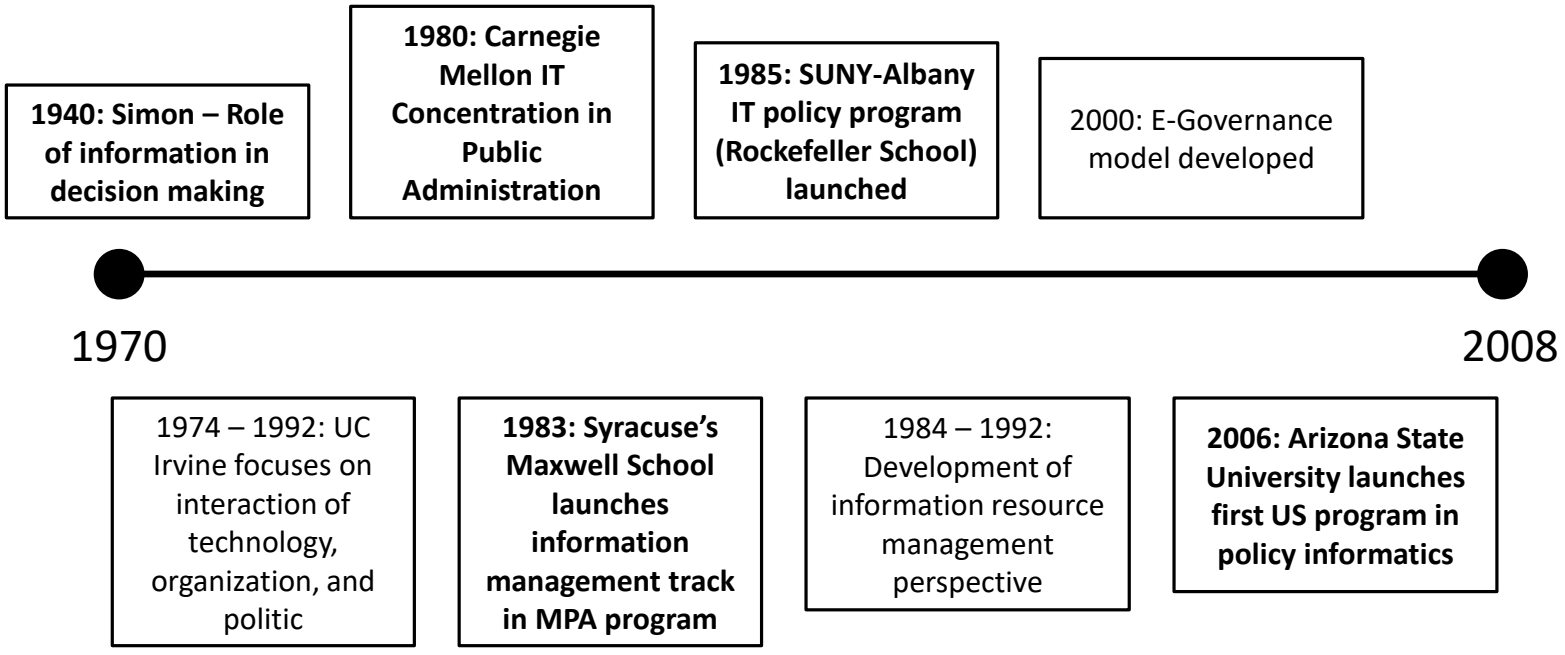
[Monitoring/ detection]
Outliers and abnormal events

[Forecasting]
Likelihood and timing

[Review/Insight]
Implications, challenges, and research agenda

2. Policy informatics

- **Origin and development**



Carnegie Mellon University (Master of Science in Public Policy & Management – Data Analytics Track)

Northwestern University (Master in Public Policy & Administration – Data Analytics for Public Policy Specialization)

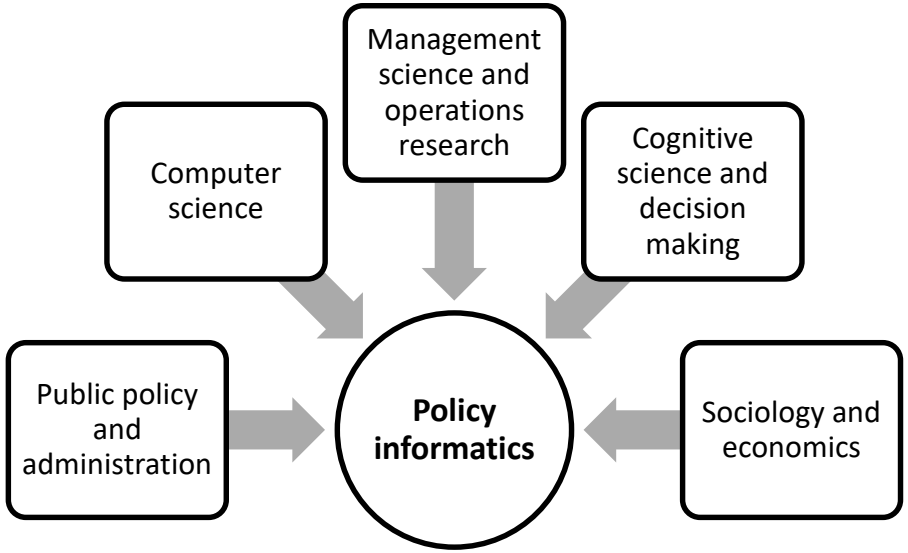
...

University of Chicago (Master of Science in Computational Analysis & Public Policy)

Kim, Y., and Johnston, E. (2008). Policy informatics v1.0, In *Minnowbrook III Conference, Lake Placid*.

2. Policy informatics

- **Assumption**
  - More intensive and creative use of information and technology → More effective policy-making processes and better policy choices
- **Definition**
  - A transdisciplinary study of how to use information and computation to understand and tackle complex problems of society
- **Relevant disciplines**



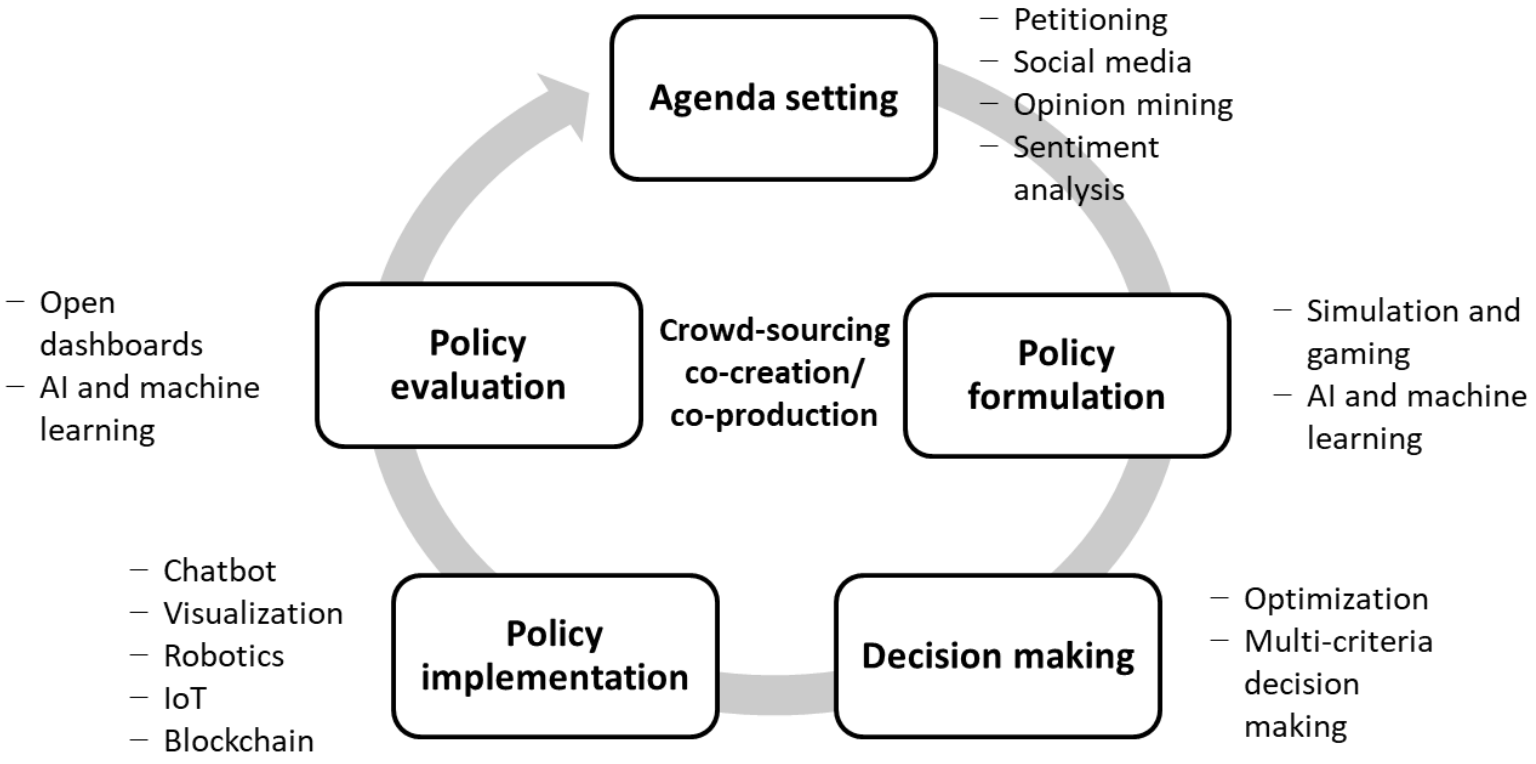
2. Policy informatics

- **Example**
  - Impact of social distancing on the spread of coronavirus

https://corona.katapult-magazin.de/

2. Policy informatics

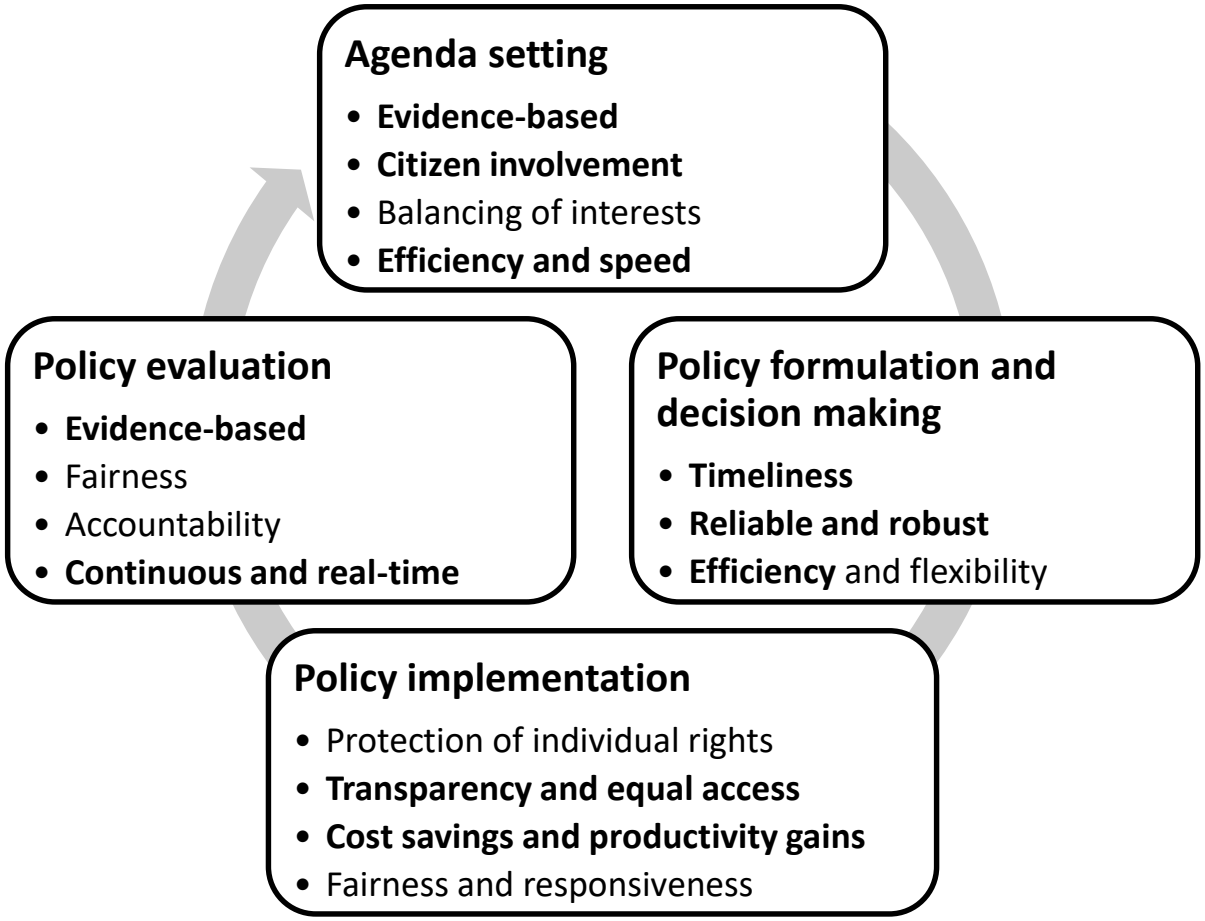
- **Current trends shaping policy informatics**
  - Disruptive technologies combined with a paradigm shift towards more citizen- and needs-driven developments in all phases of the policy life cycle



Janssen, M., Wimmer, M. A., Deljoo, A. (2015). *Policy Practice and Digital Science: Integrating Complex Systems, Social Simulation and Public Administration in Policy Research* (Vol. 10), Springer.

2. Policy informatics

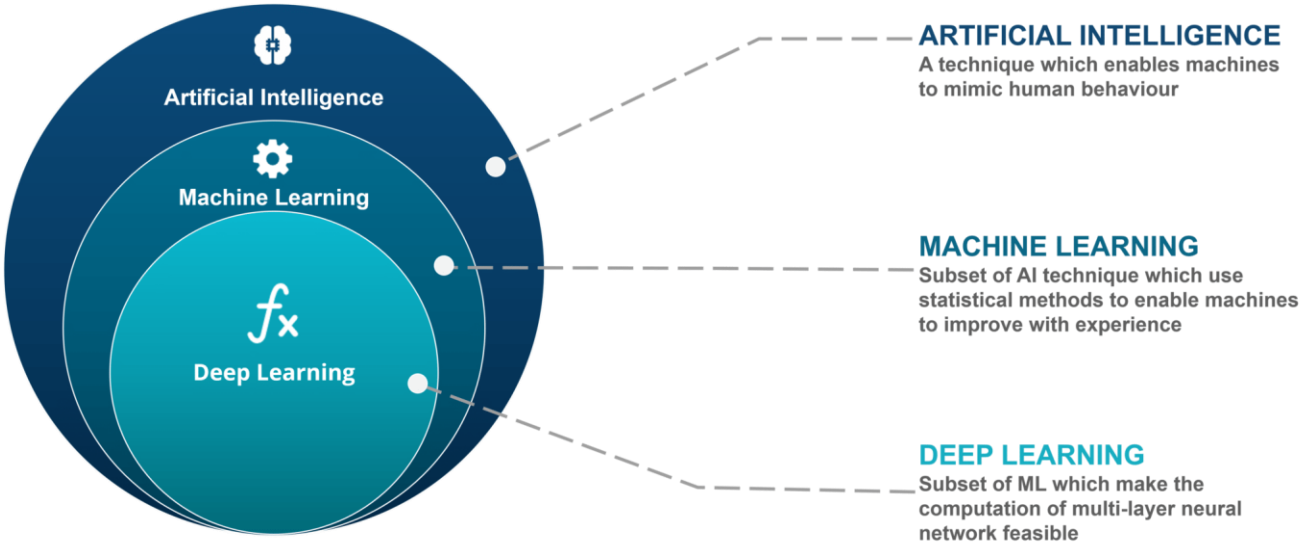
- **Public value expectations along policy life cycle**



Janssen, M., and Helbig, N. (2018). Innovating and changing the policy-cycle: Policy-makers be prepared!, *Government Information Quarterly*, *35*(4), S99-S105.

3. AI and machine learning

- **Definition of machine learning**
  - A computer program is said to **“learn”** from **experience E** with respect to some class of **tasks T** and **performance measure P**, if its performance at tasks in **T**, as measured by **P**, improves with **E**.
  - A **“machine learning algorithm”** is an algorithm that is able to **learn from data**
    - ✓ Machine learning = Learning from **E** in form of **Data**
    - ✓ **Induction over large-scale data**
- **AI vs. machine learning vs. deep learning**



3. AI and machine learning

- **Use of AI in the public sector**

분야	예시
교통	• 자율주행자동차, 셔틀(교통 체증, 사고 해소) • 항공, 해운 활용
스마트시티	• 효율적 도시 관리(지능형 교통 시스템) • CCTV를 활용한 안전사회 구현 및 법 집행(치매, 실종유아 등 찾기;과기정통부)
의료관리	• 정밀의료, 처방, 신속한 진단
사이버 보안	• 해킹 등 위험 발굴 및 대응
금융	• 보이스피싱 탐지 및 차단기술의 활용(한국 연간 6조 피해) • 신용위기 분석(한국 부동산 정보 활용) 한국은행 금리결정 도입시 활용 검토
안보	• 신병 모집 시 Chatbot 활용(美)
사법서비스	• 빅데이터 분석에 의한 판결
자연재해	• IBM, OmniEarth 캘리포니아 가뭄 해결 시도(수요 예측 등)
통계	• 빅데이터 분석에 기반한 인구통계 처리 등

정소윤, 이재호, 강정석 (2019). 인공지능(AI) 행정 도입에 따른 변화와 활성화 조건, 한국행정연구원.

II. Expert-machine collaboration for decision making

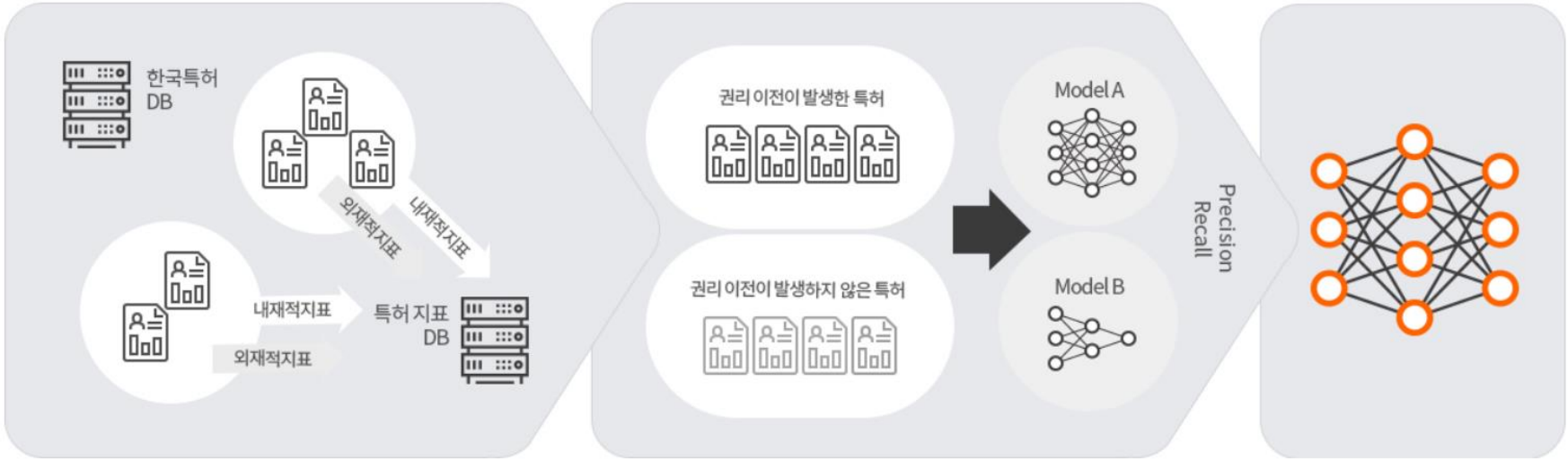
Kim, J., Lee, G., Lee, S., and **Lee, C.***. Towards expert-machine collaborations for technology valuation: An interpretable machine learning approach, Under 3rd round review at *Technological Forecasting and Social Change*.

1. Research background and problem

- **Korea Invention Promotion Association**
  - SMART3 for technology valuation

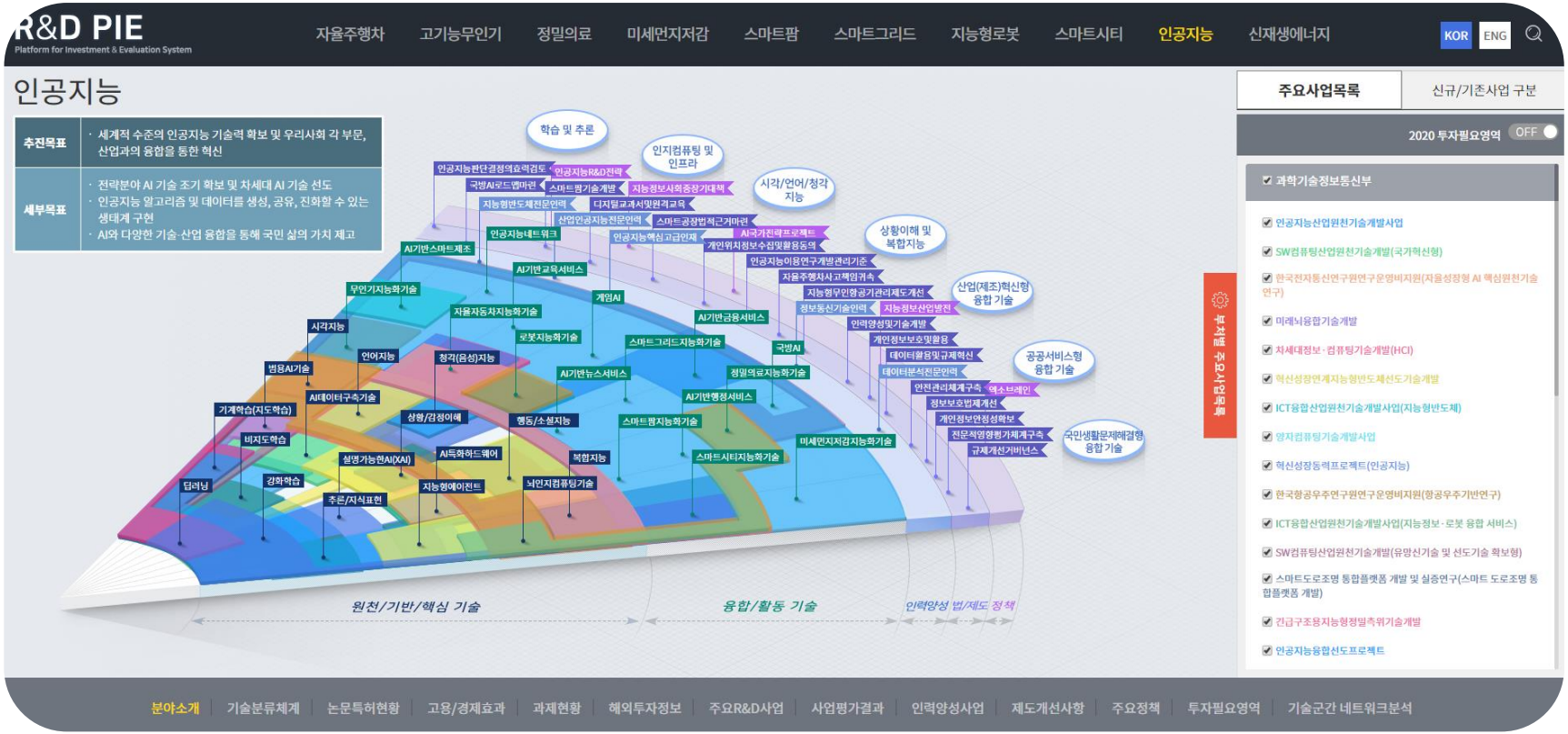
1. Research background and problem

- **Korea Technology Finance Corporation**
  - KTRS for technology valuation



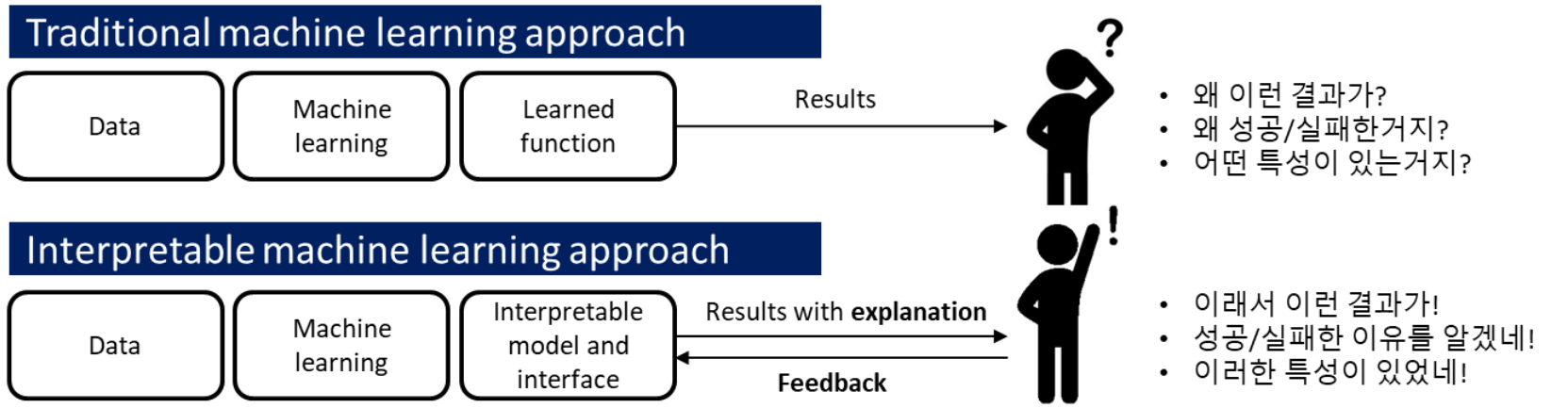
1. Research background and problem

- **Ministry of Science and ICT**
  - R&D PIE for R&D investment



1. Research background and problem

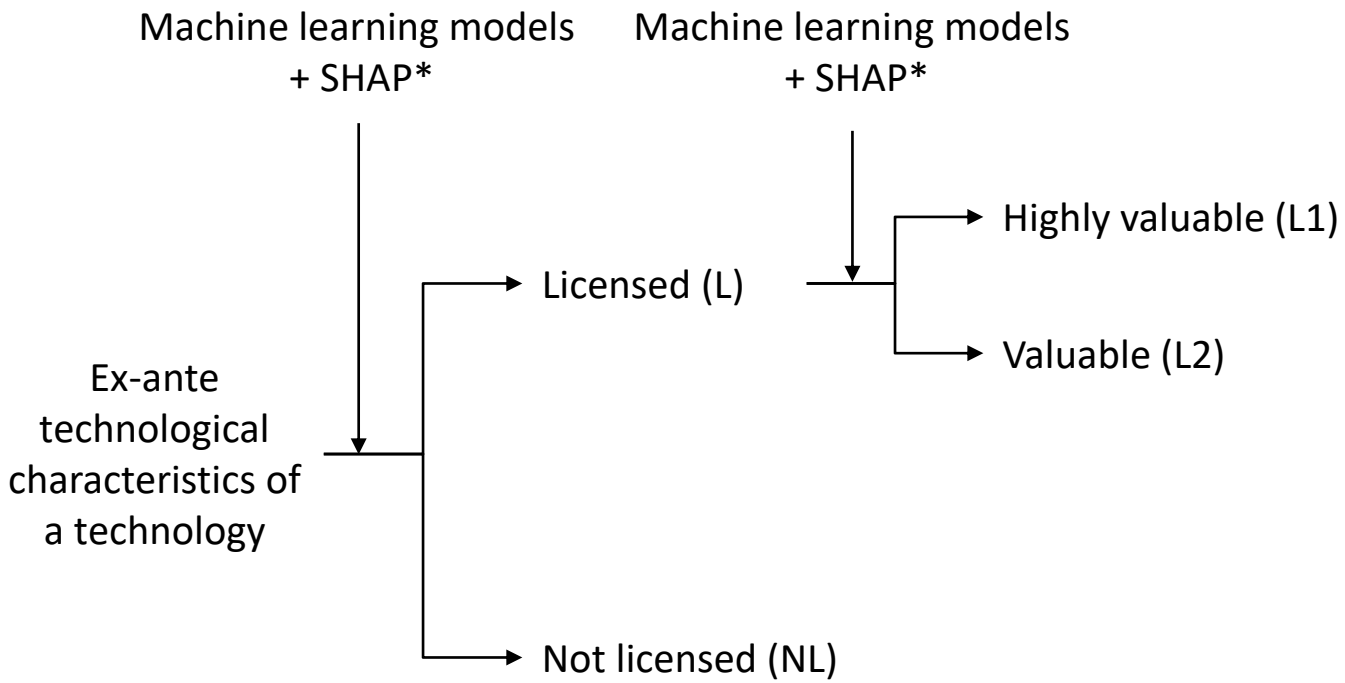
- **Motivation**
  - **High level of uncertainty and complexity** associated with technology valuation
  - Previous models relying solely on black-box models



- **Objective**
  - To develop an analytical framework for successful expert-machine collaborations for technology valuation using interpretable machine learning models

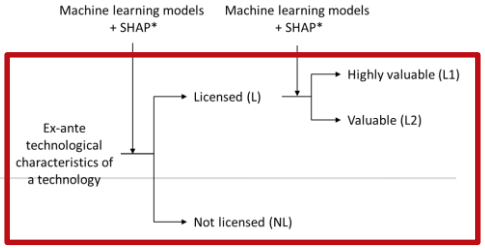
2. Data and methodology

- **High-level description of the data and machine learning models**



*: SHapley Additive exPlanation

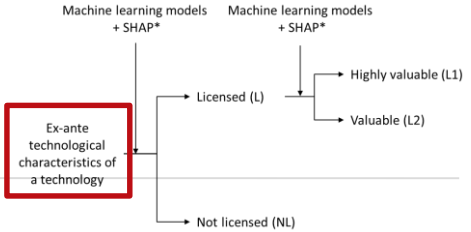
2. Data and methodology



- **Data**
  - Technology transaction database
    - ✓ A full sample of inventions that were disclosed to the Office of Technology Licensing of Stanford University from January 1970 to July 2014
    - ✓ To measure technology value and marketing activities
  - Patent database
    - ✓ To measure technological characteristics

Category	Subcategory	Economic value	Number of technologies
Licensed (L)	Highly valuable (L1)	Above $500,000	68 (4.14%)
	Valuable (L2)	0–$500,000	768 (46.74%)
Not licensed (NL)		0	807 (49.12%)
Sum			1,643 (100%)

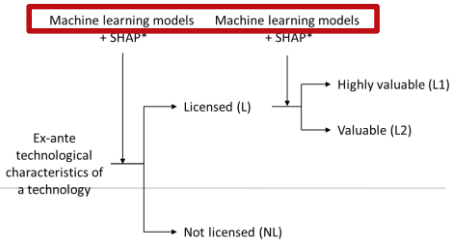
2. Data and methodology



- **Summary of the features of technological characteristics**

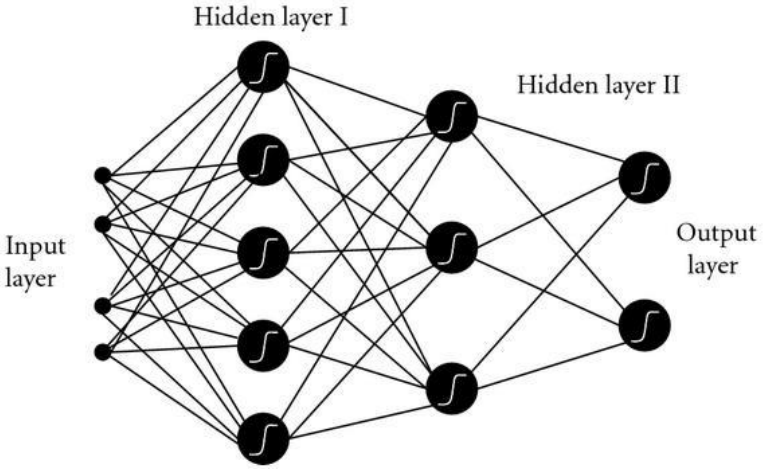
Category	Data source	Feature	Operational definition	References
Technological novelty and originality	Patent database	Technology age (TA)	The amount of time between a technology being registered in the OTL and being licensed (or the current time)	Fischer and Leidinger (2014)
		Prior knowledge (PK)	Number of backward citations of the patents for a technology	Harhoff et al. (2003)
		Scientific knowledge (SK)	Number of non-patent literature references of the patents for a technology	Callaert et al. (2006)
		Technology cycle time (TCT)	Median age of cited patents	Bierly and Chakrabarti (1996)
		Main class-level originality (MCO)	Herfindahl index on classes of cited patents	Bessen (2008); Jaffe and Trajtenberg (2002)
		Subclass-level originality (SCO)	Herfindahl index on mainline subclasses of cited patents	
		Examination time (ET)	Time difference between the first patent publication and the patent application	Higham et al. (2021)
Technological scope	Patent database	Patent count (PC)	Number of patents for a technology	Hirschey and Richardson (2004)
		Main class count (MCC)	Number of main classes of the patents for a technology	Lerner (1994)
		Subclass count (SCC)	Number of mainline subclasses of the patents for a technology	
	Technology transaction database	Bio science relevance (Bio)	1 if a technology is related to bio science, otherwise 0	-
Technological superiority	Patent database	Independent claims (IC)	Number of independent claims of patents for a technology	Lanjouw and Schankerman (2001)
		Dependent claims (DC)	Number of dependent claims of patents for a technology	
	Technology transaction database	Federal government fund (FGF)	1 if technology development is funded by federal governments, otherwise 0	Corredoira et al. (2018)
		Edison awards winner (EAW)	1 if a technology wins the Edison awards, otherwise 0	-
Market coverage	Patent database	Patent family (PF)	Number of patents registered in multiple countries with the coverage of the same invention	Gullec and Potterie (2000)
	Technology transaction database	Application area (AA)	Number of potential application areas of a technology	-
Development efforts and capabilities	Patent database	Human resources (HR)	Number of inventors of the patents for a technology	Ma and Lee (2008)
		Collaboration (Col)	1 if patents for a technology have more than one assignee, otherwise 0	Ma and Lee (2008)
Sponsorship and marketing	Technology transaction database	Sponsors (Spon)	Number of sponsors for technology development	Wright et al. (2014)
		Recipients (Recip)	Number of marketing recipients	-

2. Data and methodology

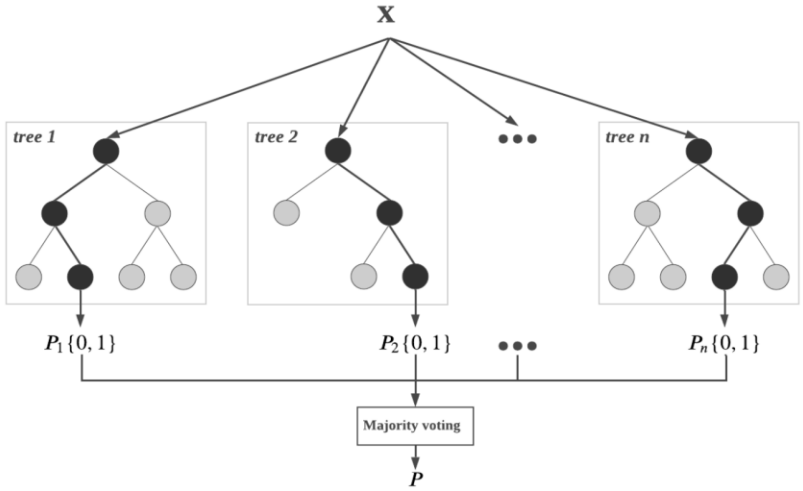


- **Methodology**
  - Five machine learning models for assessing the economic value of technologies
    - ✓ Multi-layer perceptron (MLP)
    - ✓ Support vector machine (SVM)
    - ✓ Factorization machine (FM)
    - ✓ Random forest (RF)
    - ✓ Extreme gradient boosting (XGBoost)

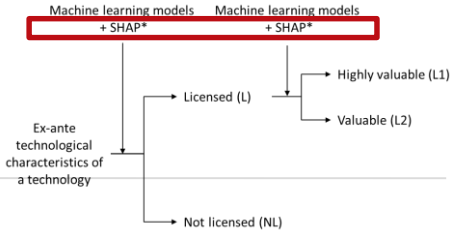
<MLP>



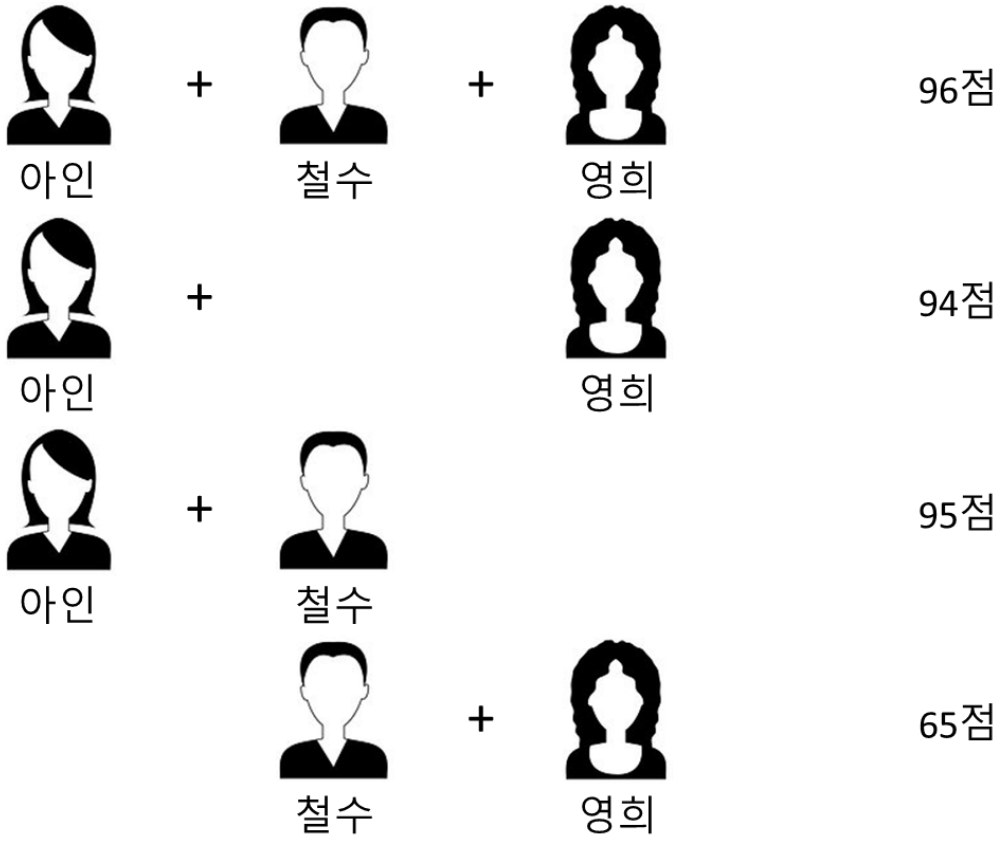
<RF>



2. Data and methodology



- **Methodology**
  - SHAP for interpreting the models' mechanisms and behaviors



3. Empirical analysis and results

- **Results of technology valuation using machine learning models**

Technology ID	Technological characteristics					(Actual) Economic value	(Predicted) Economic value									
	TA	PK	…	Spon	Recip		MLP		SVM		FM		RF		XGBoost	
00-003	20	2	…	2	368		L	L2	L	L2	L	L2	L	L2	L	L2
00-009	191	21	…	1	1367		NL	-	NL	-	NL	-	NL	-	NL	-
00-010	206	5	…	1	7		NL	-	NL	-	NL	-	NL	-	NL	-
00-045	−1	44.5	…	2	131		L	L2	L	L1	L	L2	L	L2	L	L1
…	…	…	…	…	…		…	…	…	…	…	…	…	…	…	…
02-164	6	12	…	1	6		L	L2	L	L1	L	L2	L	L2	L	L2
02-166	−3	6	…	1	0		L	L2	L	L2	L	L2	L	L2	L	L2
02-170	206	4	…	1	2		NL	-	NL	-	NL	-	NL	-	NL	-
02-181	50	2		1	78		L	L2	L	L1	L	L2	L	L2	L	L1
…	…	…	…	…	…		…	…	…	…	…	…	…	…	…	…
99-220	154	18	…	1	151		NL	L2	L	L2	NL	L2	NL	L2	NL	L2
99-231	−3	7.33	…	5	3		L	L2	L	L2	L	L1	L	L2	L	L1
99-236	202	2	…	1	158		NL	-	NL	-	L	L2	NL	-	NL	-

3. Empirical analysis and results

- **Results of technology valuation using machine learning models**

Technology ID	Technological characteristics					(Actual) Economic value		(Predicted) Economic value									
	TA	PK	…	Spon	Recip			MLP		SVM		FM		RF		XGBoost	
00-003	20	2	…	2	368	L	L2	L	L2	L	L2	L	L2	L	L2	L	L2
00-009	191	21	…	1	1367	NL	-	NL	-	NL	-	NL	-	NL	-	NL	-
00-010	206	5	…	1	7	NL	-	NL	-	NL	-	NL	-	NL	-	NL	-
00-045	−1	44.5	…	2	131	L	L1	L	L2	L	L1	L	L2	L	L2	L	L1
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
02-164	6	12	…	1	6	L	L1	L	L2	L	L1	L	L2	L	L2	L	L2
02-166	−3	6	…	1	0	L	L2	L	L2	L	L2	L	L2	L	L2	L	L2
02-170	206	4	…	1	2	NL	-	NL	-	NL	-	NL	-	NL	-	NL	-
02-181	50	2		1	78	L	L1	L	L2	L	L1	L	L2	L	L2	L	L1
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
99-220	154	18	…	1	151	L	L2	NL	L2	L	L2	NL	L2	NL	L2	NL	L2
99-231	−3	7.33	…	5	3	L	L2	L	L2	L	L2	L	L1	L	L2	L	L1
99-236	202	2	…	1	158	NL	-	NL	-	NL	-	L	L2	NL	-	NL	-

3. Empirical analysis and results

- **Summary of performance evaluation**

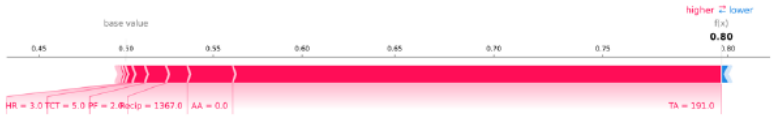
Model	Level of analysis	Accuracy	Precision	Recall	Specificity	F1 score
MLP	L vs. NL	0.93	**0.97**	0.90	**0.97**	0.93
	L1 vs. L2	0.90	0.22	0.10	0.97	0.14
SVM	L vs. NL	0.56	0.54	**1.00**	0.10	0.70
	L1 vs. L2	**0.92**	0.00	0.00	**1.00**	0.00
FM	L vs. NL	0.90	0.88	0.93	0.87	0.91
	L1 vs. L2	0.84	0.21	0.34	0.89	0.26
RF	L vs. NL	**0.95**	**0.97**	0.93	**0.97**	**0.95**
	L1 vs. L2	**0.92**	**0.33**	0.03	0.99	0.05
XGBoost	L vs. NL	0.94	**0.97**	0.91	**0.97**	0.94
	L1 vs. L2	0.73	0.20	**0.75**	0.73	**0.31**

3. Empirical analysis and results

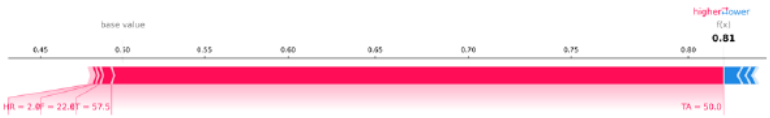
- **SHAP values for technology valuation**

(a) SHAP values for classifying 00-009 (actual NL) as L

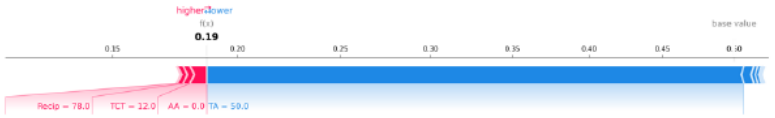
(b) SHAP values for classifying 00-009 (actual NL) as NL



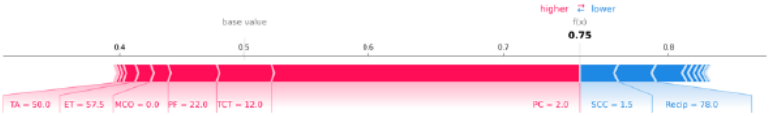
(c) SHAP values for classifying 02-181 (actual L) as L



(d) SHAP values for classifying 02-181 (actual L) as NL

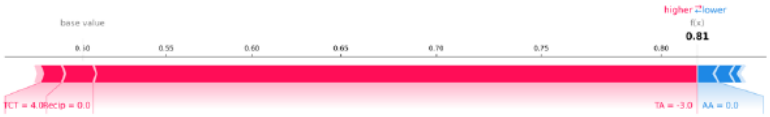


(e) SHAP values for classifying 02-181 (actual L1) as L1

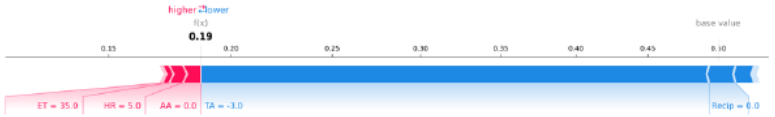


(f) SHAP values for classifying 02-181 (actual L1) as L2

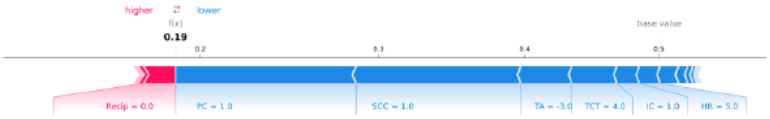
(g) SHAP values for classifying 02-166 (actual L) as L



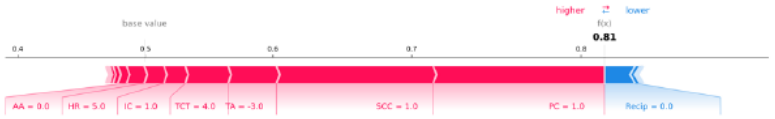
(h) SHAP values for classifying 02-166 (actual L) as NL



(i) SHAP values for classifying 02-166 (actual L2) as L1

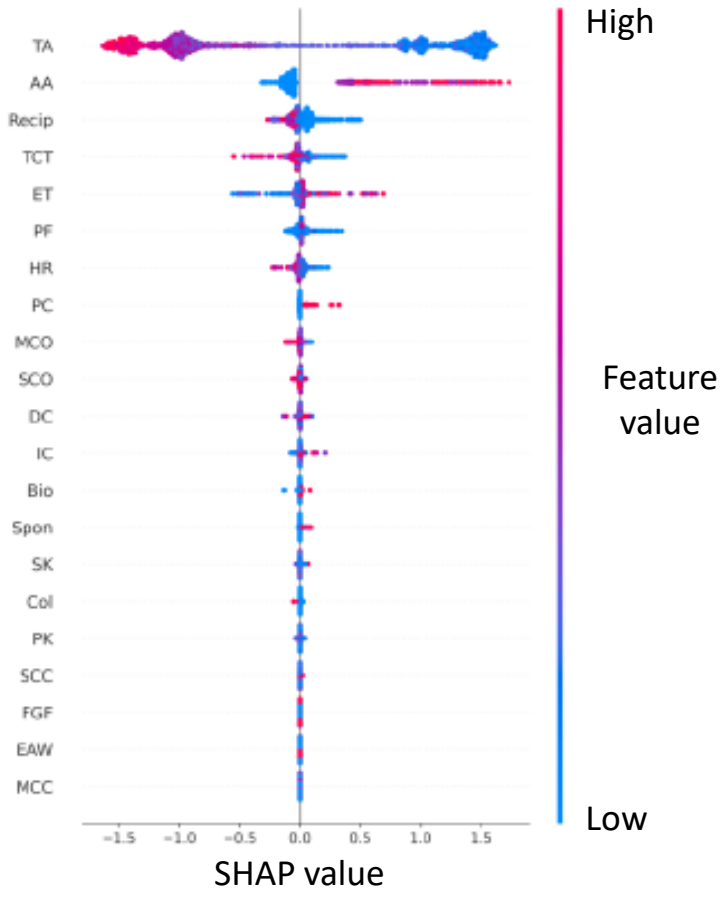


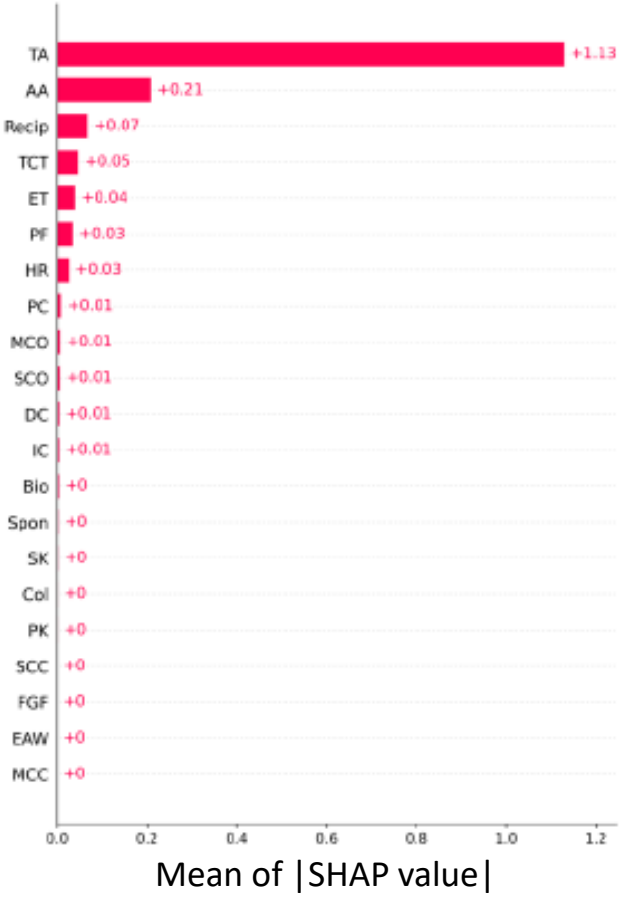
(j) SHAP values for classifying 02-166 (actual L2) as L2



3. Empirical analysis and results

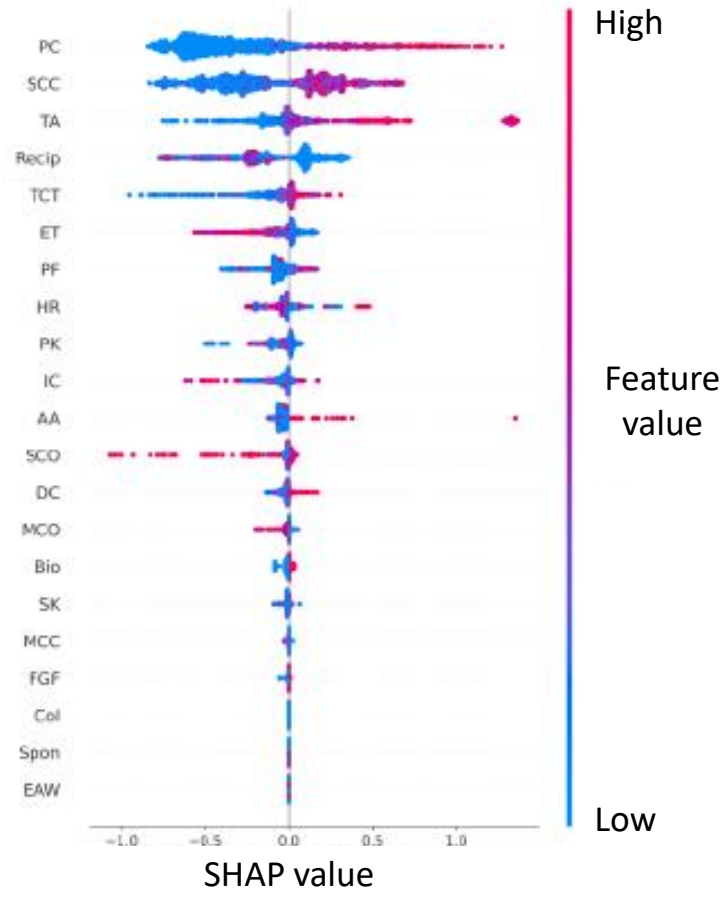
- **Feature importance**
  - Summary plot (L vs. NL)
  - Bar plot (L vs. NL)

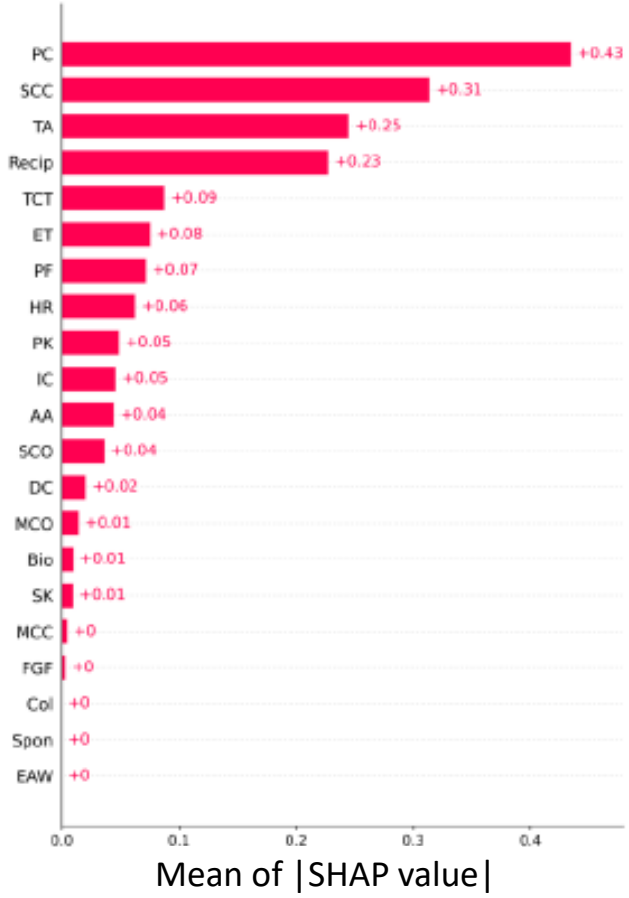




3. Empirical analysis and results

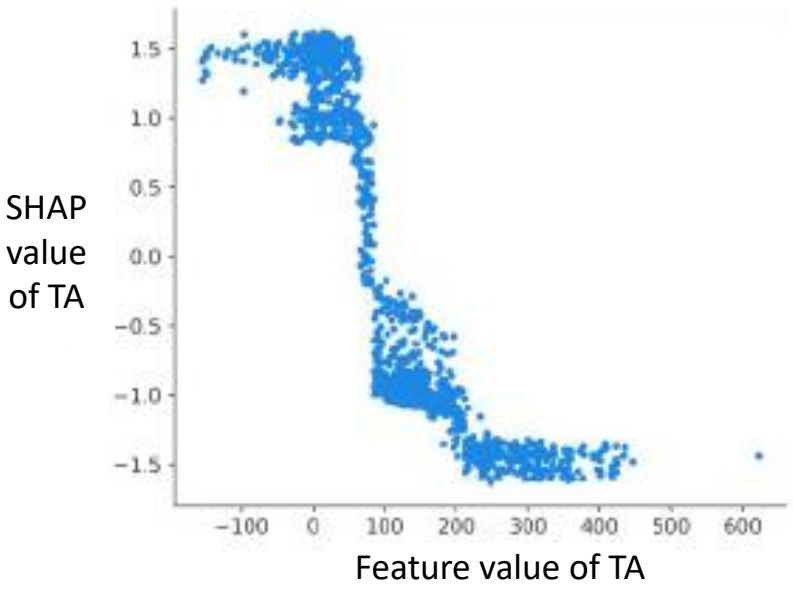
- **Feature importance**
  - Summary plot (L1 vs. L2)
  - Bar plot (L1 vs. L2)

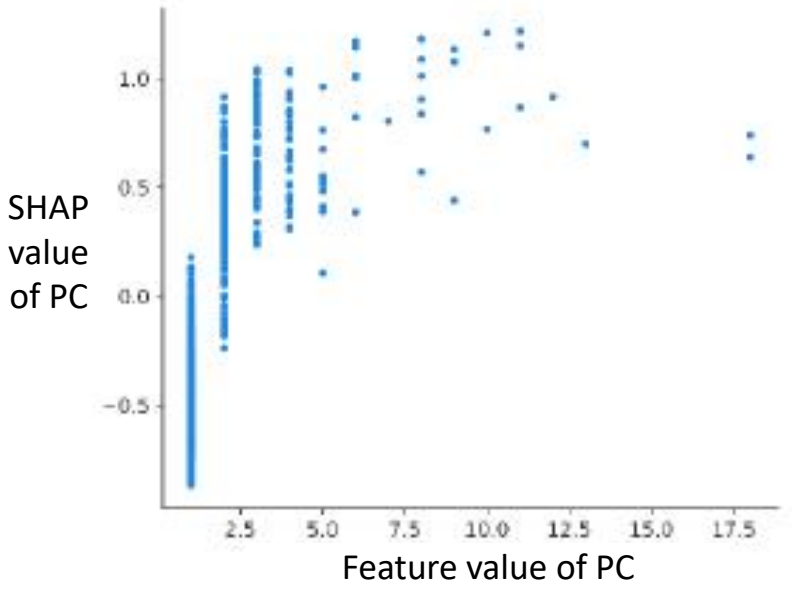




3. Empirical analysis and results

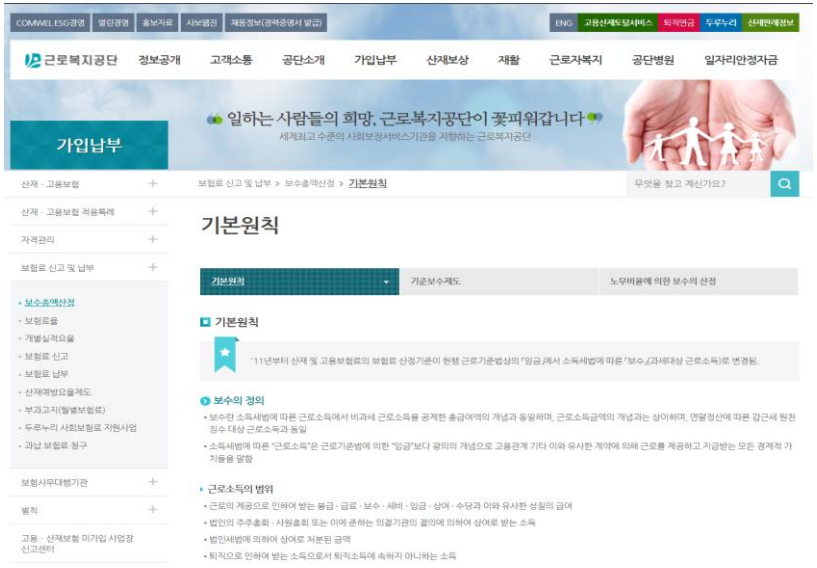
- **Feature dependence and interaction**
  - Dependence plot of TA (L vs. NL)
  - Dependence plot of PC (L1 vs. L2)

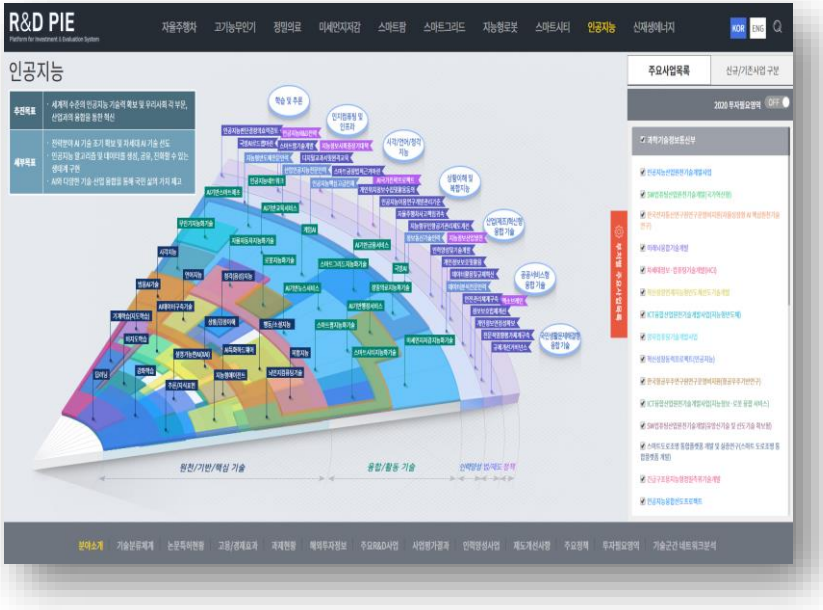




4. Potential applications

- **Decision making under the high level of complexity and uncertainty**
  - Inspection and audit
  - Resource allocation



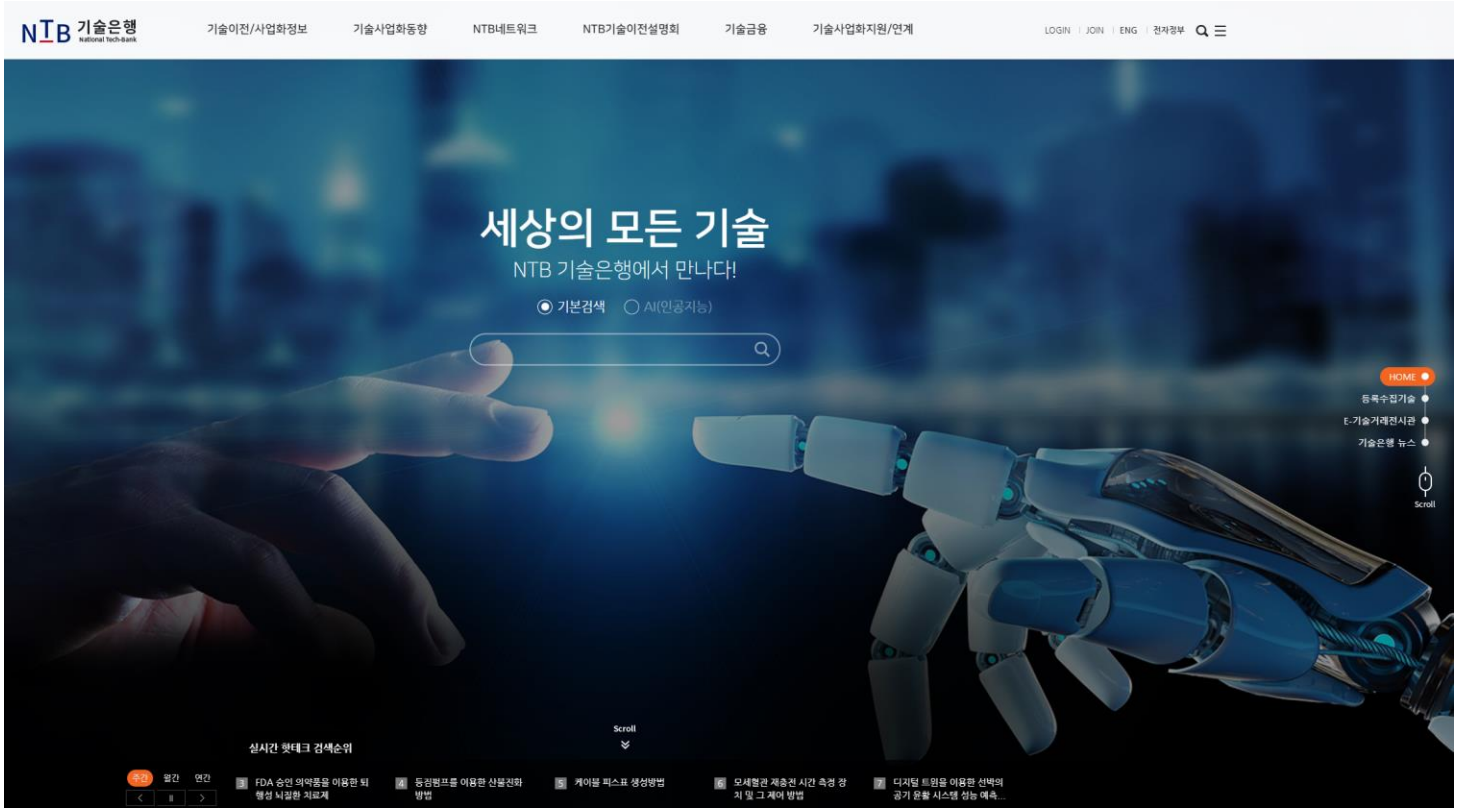


III. Mismatch between the supply and demand

Lee, G., Lee, S., Kim, C., and **Lee, C.***. Inventor-licensee matchmaking in university technology licensing contexts: Linking technological functions and business requirements using representation learning, In preparation for *Technological Forecasting and Social Change*.

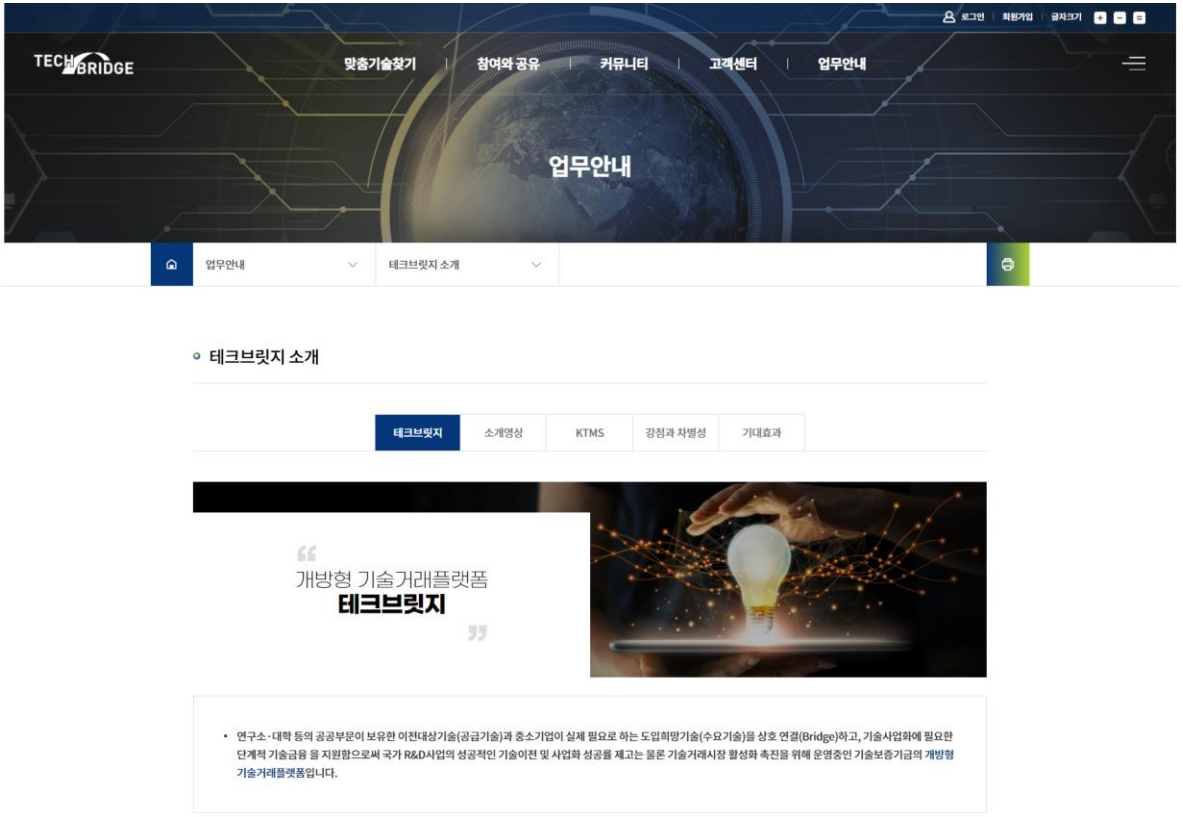
1. Research background and problem

- **Ministry of Trade, Industry and Energy**
  - National Tech-bank



1. Research background and problem

- **Korea Technology Finance Corporation**
  - TechBridge



1. Research background and problem

- **Background**
  - Increasing importance of university-industry-government interactions
  - Necessities of developing systematic approaches to matchmaking in technology licensing contexts
- **Motivation**
  - Disparity between technical and business languages

  → 기업은 생산 현장에서 발생하는 안전사고 감소 방법을 요구

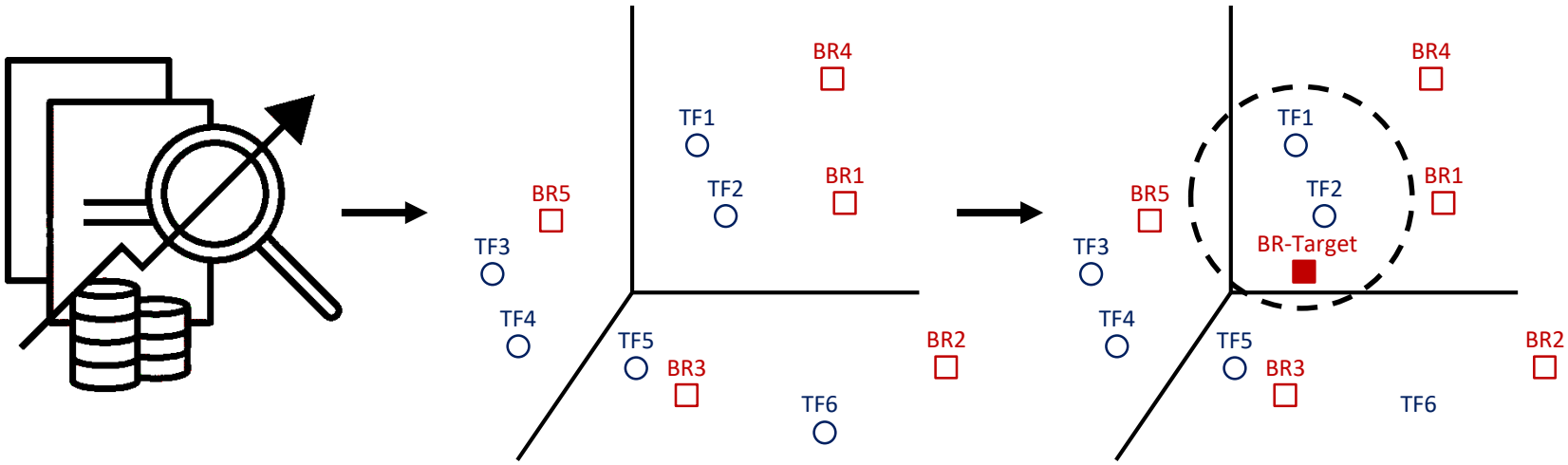
  → 모션 탐지 기술을 보유한 연구자 추천
  - No perfect match between the technological functions and business requirements

  → 기업은 참외를 5mm로 깍는 방법을 요구

  → 사과를 3mm로 깍는 기술을 보유한 연구자 추천
- **Objective**
  - To develop an analytical framework for inventor-licensee matchmaking in university technology licensing contexts based on representation learning

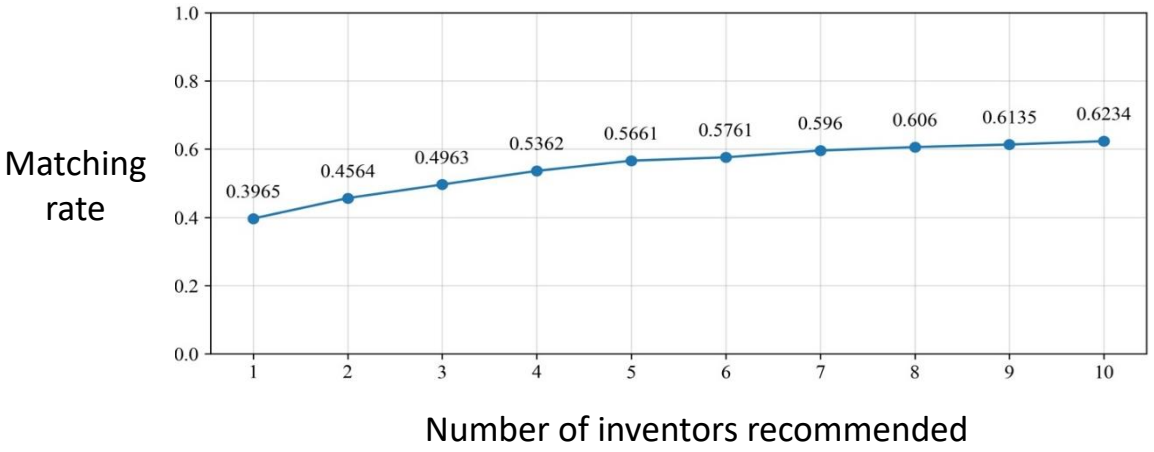
2. Data and methodology

- **Data**
  - Patent, publication, and project databases
    - ✓ To measure technological functions that inventors can offer
  - Business requirement database
    - ✓ To measure business requirements that licensees demand
  - Technology and know-how licensing databases
    - ✓ To assess the performance and utility of the proposed approach
- **Methodology: Representation learning (e.g., fastText and BERT)**

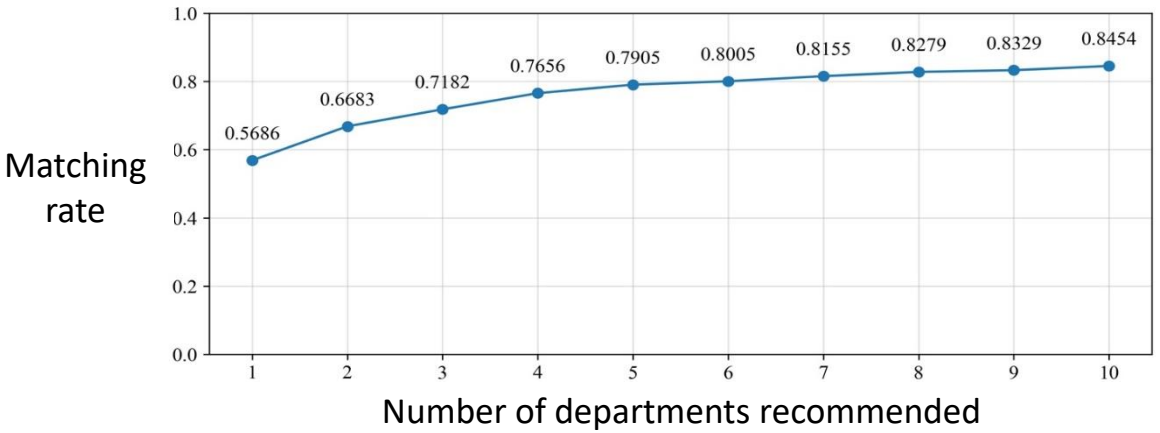


3. Empirical analysis and results

- **Matching rate at the inventor level**

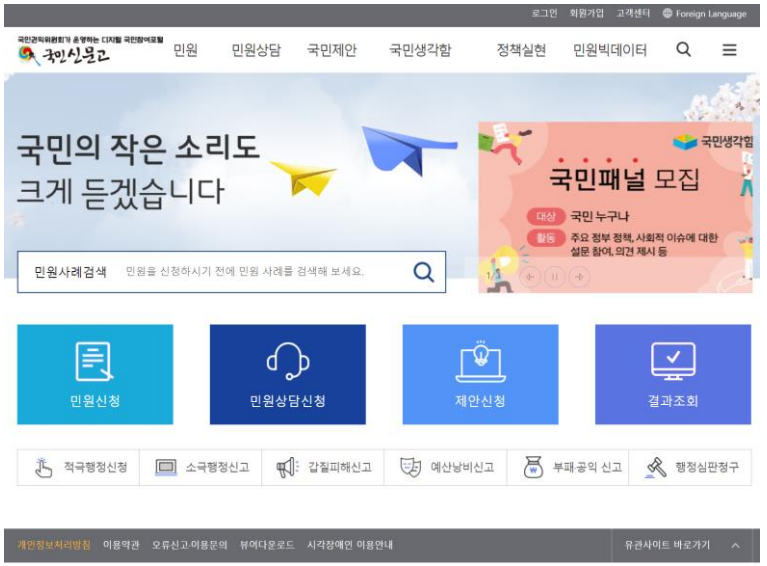


- **Matching rate at the department level**



4. Potential applications

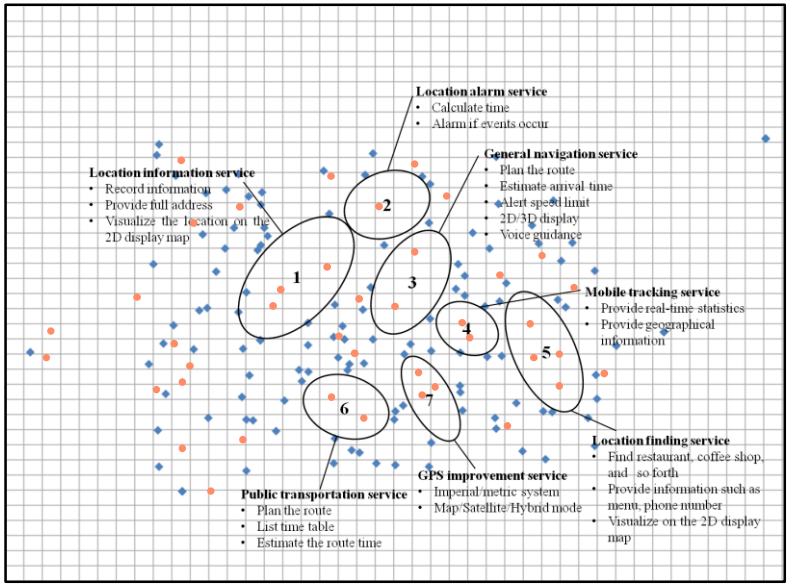
- **Various types of mismatches between the supply and demand**
  - Labor market mismatch
  - Complaint-solution mismatch

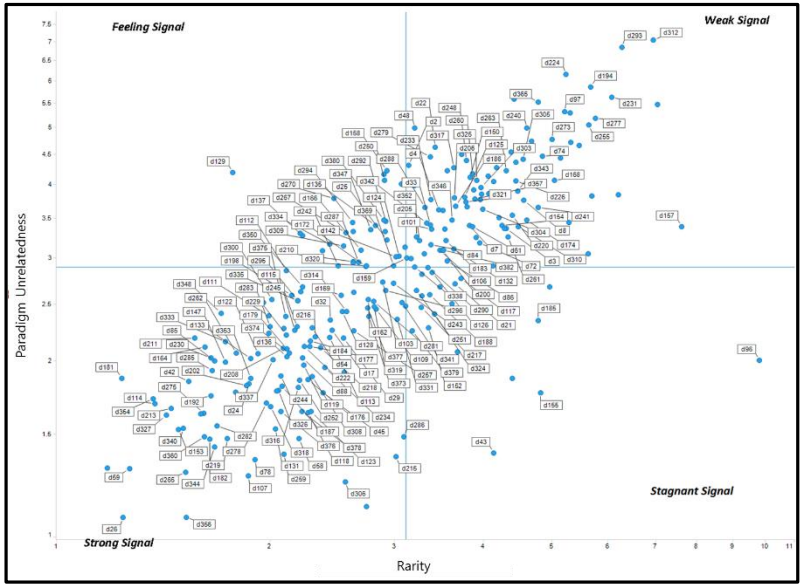


IV. Opinion mining, sentiment analysis, and simulation

1. Opinion mining for agenda setting

- **Identifying new service opportunities from large scale documents**





Lee, C., and Lee, H.* (2015). Novelty-focused document mapping to identify new service opportunities, *Service Industries Journal*, 35(6), 345–361.

Kim, J., and **Lee, C.*** (2017). Novelty-focused weak signal detection in futuristic data: Assessing the rarity and paradigm unrelatedness of signals, *Technological Forecasting and Social Change,* 120, 59–76.

2. Sentiment analysis for agenda setting and policy evaluation

- **Assessing service quality using customer reviews**

Review 1

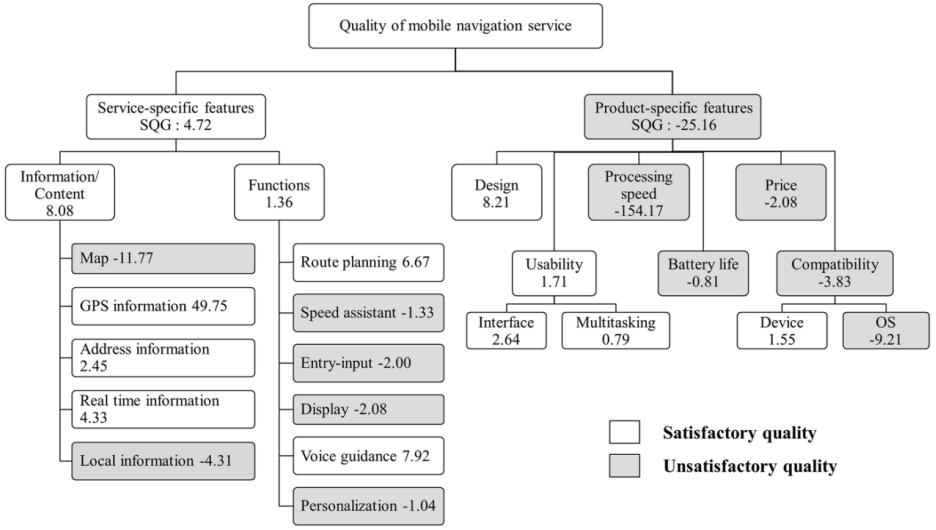
Overall rating: 5
Location is *bad*
View is *attractive*

Review 2

Overall rating: 3
Location is *bad*
View is *good*

Review	OR	SFrK (Service feature-related keyword)		
		Location	View	Food
		CS_{i1}	CS_{i2}	CS_{i3}
CO_1	5	-1	1	0
CO_2	3	-1	1	0
…	…	…	…	…

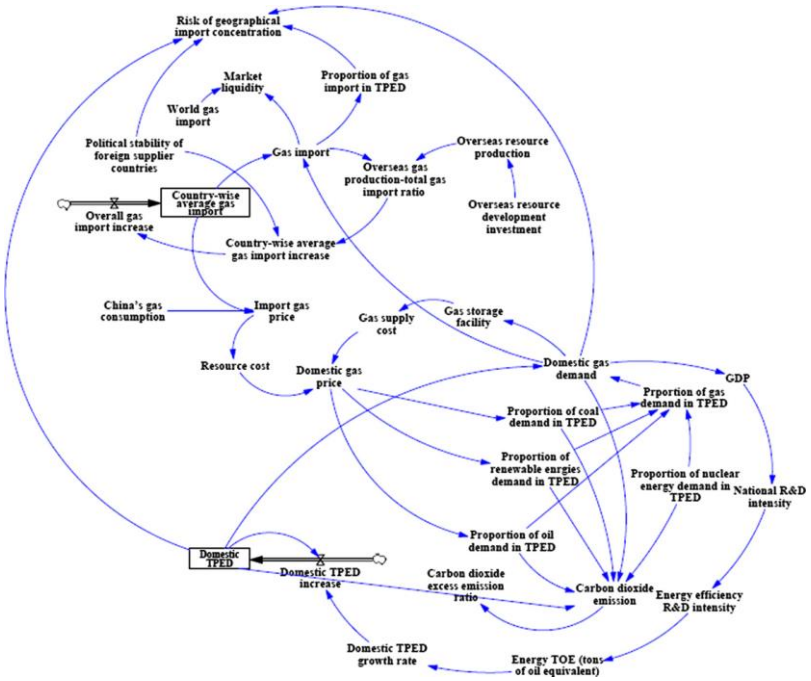
- *bad*: a negative StrK (sentiment-related keyword)
- *good*, *attractive*: positive StrKs

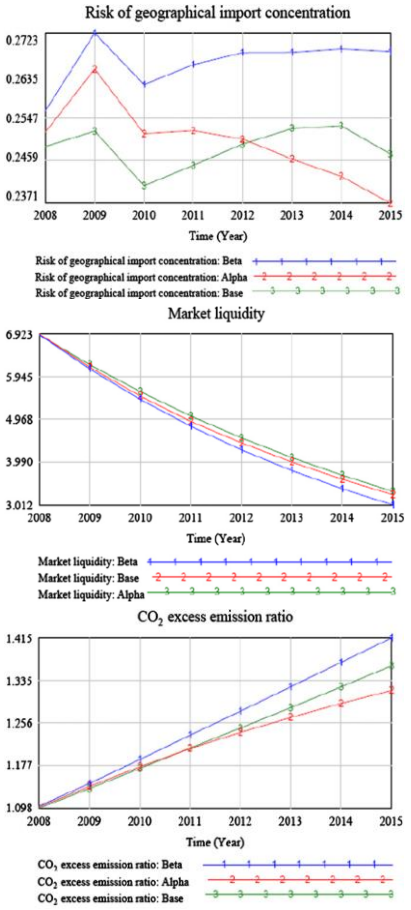


Song, B., **Lee, C.**, Yoon, B., and Park, Y.* (2016). Diagnosing service quality using customer reviews: An index approach based on sentiment and gap analyses, *Service Business*, 10(4), 775–798.

3. Simulation for policy evaluation

- **Energy security management model using system dynamics**





Shin, J., Shin, W., and **Lee, C.*** (2013). An energy security management model using quality function deployment and system dynamics, *Energy Policy*, 54, 72–86.

4. Chatbot for policy implementation

- **Sentence BERT* practice**

```
In [8]: return_answer('안녕~ 반가워!')
Out[8]: '안녕하세요.'

In [9]: return_answer('너 말 잘한다')
Out[9]: '그런 사람이 있으면 저 좀 소개시켜주세요.'

In [10]: return_answer('나랑 커피마실까?')
Out[10]: '좋은 시간 보내시길 바라요.'

In [11]: return_answer('아인이는 너무 귀여워')
Out[11]: '귀여운 사람이 말해서 그래요.'

In [12]: return_answer('나는 2022년도 3월부터 서강대 MOT 대학원에서 공부를 시작해~')
Out[12]: '꿈에 도전하는 건 좋은 거라고 들었어요.'

In [13]: return_answer('의사결정 지원을 위한 인공지능의 활용 수업은 어려울까?')
Out[13]: '나한테 맞는 공부 방법 찾는 게 시급하네요.'

In [14]: return_answer('열심히 하면 잘할수 있겠지?')
Out[14]: '인내의 시간이 필요할 거예요.'

In [15]: return_answer('A+ 받자! 화이팅!')
Out[15]: '성공을 기원합니다.'
```

*: Bidirectional Encoder Representations from Transformers

IV. Conclusion

1. Challenges and paths to the next stage

- **Development of guidelines on using AI in the public sector**
  - Understanding AI
    - ✓ How AI can help
    - ✓ What AI cannot do
  - Assessing if AI is the right solution
    - ✓ Public value, ethics, fairness, safety, and privacy
  - Planning and preparing for AI-based public service development
    - ✓ Robust data governance principles (quality, standards, and integration)

- **Humanistic social science-centered X+AI rather than AI+X approach**

Thank you

Graduate School of Management of Technology
Sogang University

Changyong Lee, Ph.D.
Email: changyong@sogang.ac.kr