
KAIST 데이터사이언스 대학원 입학 설명회

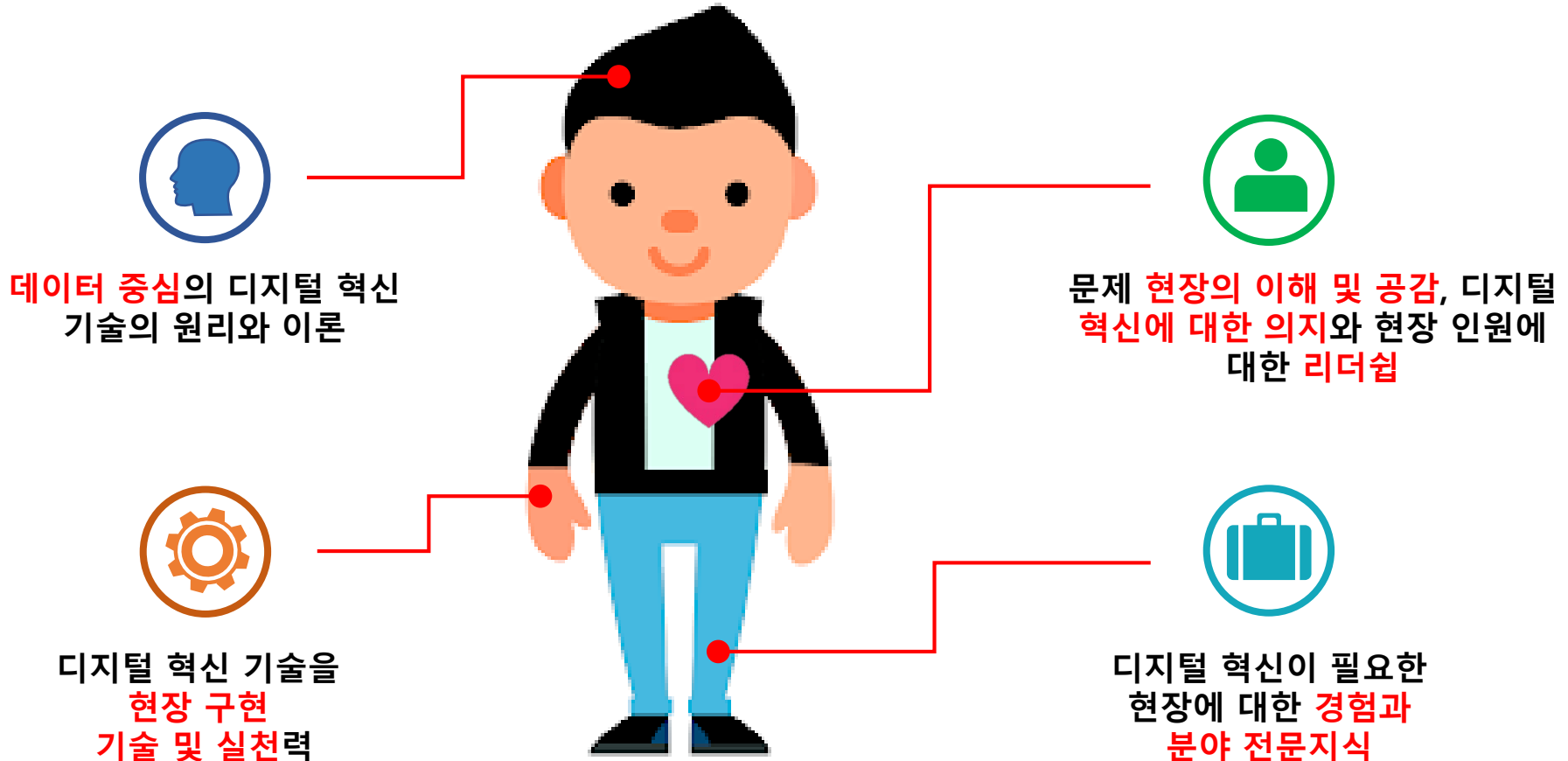
KAIST GSDS

KAIST 데이터사이언스 대학원
비전 및 인재상

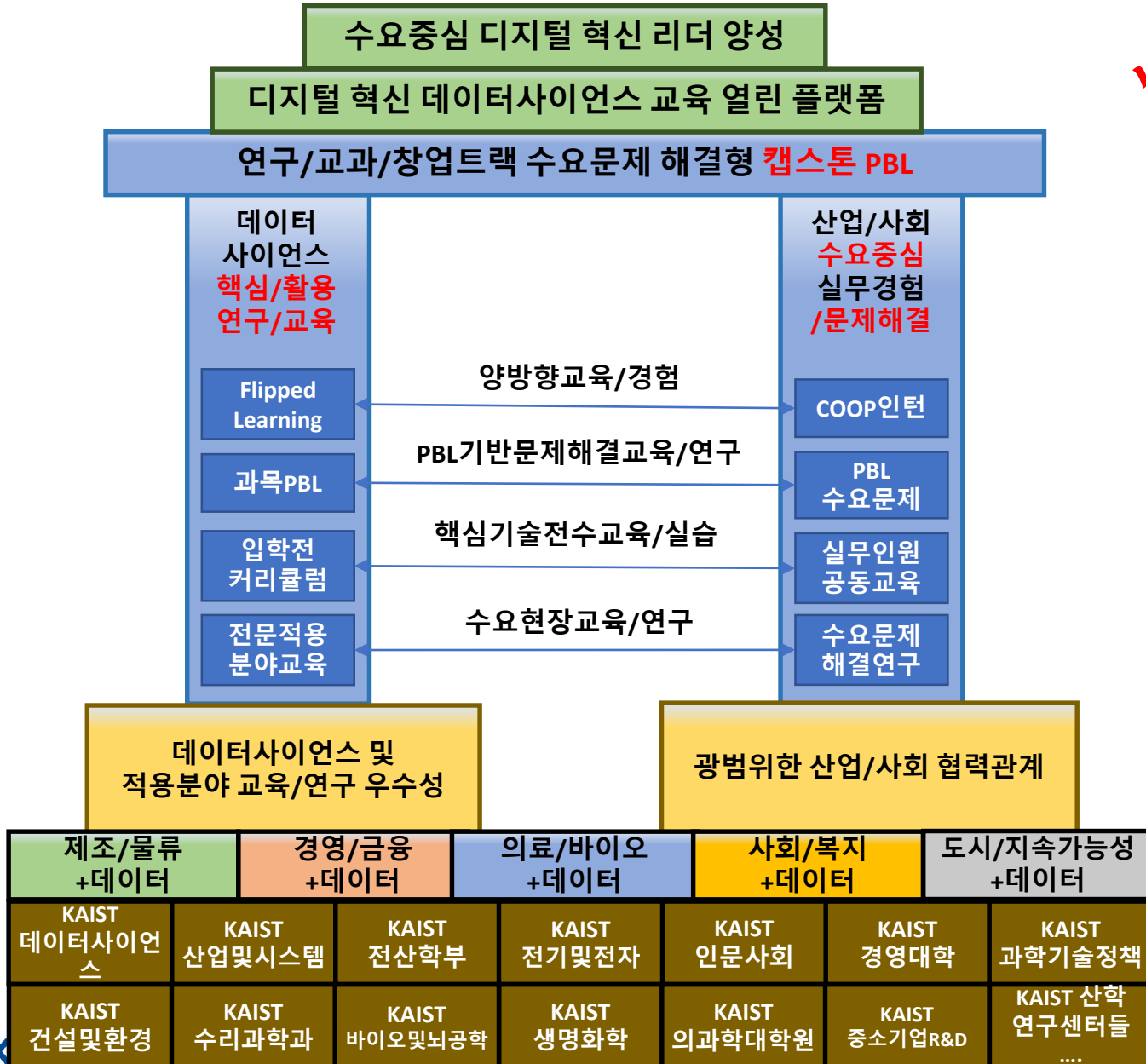
KAIST 데이터사이언스 대학원의 비전



“산업/사회 디지털 혁신의 **열린 리더**”



KAIST 데이터사이언스 대학원의 교육방침



✓ GSDS의 두 기둥

- 데이터사이언스 핵심/활용 연구/교육**
 - 최고 수준의 데이터사이언스 및 유관방법론 연구 및 교육
- 산업/사회 수요중심 문제해결**
 - 산업과 사회의 실무의 문제를 이해할 수 있고, 해결할 수 있는 관심과 의지

데이터사이언스 **방법론/응용**에 대한 최고 수준의 교육 및 도전

참여 교수진

✓ 51명 참여 교수진

- 데이터사이언스 방법론 : 15명
- 제조: 5명
- 바이오/헬스케어/뇌과학 :10명
- 금융/경영/경제 : 5명
- 인문/사회/복지/법률: 6명
- 도시/교통 : 7명
- 대기/해양/환경/지속가능: 3명

✓ 다양한 박사 전공

- 산업공학, 기계공학
- 도시공학, 환경공학
- 전산학, 통계학
- 경영과학, 정보시스템
- 의학, 법학, 사회학, 복지학

데이터사이언스 방법론



제조



금융, 경영, 경제



대기, 해양, 환경, 지속가능



총 7개 분야의 다양한 GSDS 참여 교수진 (51명)

헬스케어 바이오 뇌과학



인문, 사회, 복지, 법률



도시, 교통



참여인력의 우수성 (데이터사이언스 방법론)

반도체 제조 공정, 다양한 불량 칩 사례 늘어

김희영 (통계적방법론)

문일철 (인공지능, 추천시스템, 확률모델링)

이재길 (빅데이터, 데이터마이닝, 클라우드 시스템)

이의진 (HCI, IoT, 센서데이터)

참여인력의 우수성 (적용분야)

✓ 제조/물류

[비즈니스 리뷰 플러스] 장영재 교수 '국내 제조업의 디지털 대변혁' 특강
 © 유영 기자 | © 승인 2021.04.14 09:30



장영재 교수 '국내 제조업의 디지털 대변혁' (86 비즈니스 리뷰 플러스)

시너텍 주도 카이스트 혁신센터, AI 자동물류 상용화 앞둔다
 개소식 열고 R&D 본격 착수... "실용 공중 수송 데이터로 스마트팩토리의 발전 기대"

장영재 기자 | 승인 2020.08.18 오후 5:52:54

시너텍은 신성(AI) 카이스트와 주도해 설립한 첨단 제조기술 혁신센터가 반도체 자동화 물류 인공지능(AI) 알고리즘의 적용 적용과 상용화를 본격 시작한다.

카이스트는 최근 대한성대대학교에 위치한 공학관 캠퍼스에서 '첨단 제조기술 혁신센터'의 개소식을 올렸다. 개소식에는 신성 카이스트 총장, 김우창 시너텍 회장 외에 다수의 관계자가 참석했다.

첨단 제조기술 혁신센터는 카이스트 산업-시스템 공학과와 기계공학과가 공동으로 설립한 디지털 연구개발(R&D) 시설이다. 첨단 Over-Head Transfer와 AGV(Auto Guided Vehicle)를 실제 반도체 생산 현장과 동일하게 구축했다. 앞으로 반도체 자동화 물류 AI 알고리즘의 적용 적용, 자동화 물류 전송 연구, 시 기반 제어 시스템 개발 등을 진행할 예정이다.

패션 분류, 혼잡 피하는 기계 등... "AI로 산업 문제 해결"
 A 한복 기자 | 승인 2019.02.14 14:38

인공지능 연구자 모임 'AI 프렌즈', 제3차 정기세미나 13일 개최
장영재 KAIST 교수-신경호 기계연 박사, 산업용 AI 연구 소개

미국 브랜디에서 만든 '수베어' 종이와 '신상물류'가 전국 수백 개 매장으로 줄고 준비 중이다. 이때 인공지능 기술은 한 매장에 비슷한 색상이나 디자인의 옷이 풀리지 않도록 수확적으로 계산해 판매하는 역할을 한다.

반도체 공장의 천정에는 웨어비를 살아쓰는 혁신 역할을 하는 기계들이 들어간다. 공정은 기계가 수반 되어야만 고품재질은 찾아볼 수 있다. 이 기계에는 스스로 교정하여 원할한 일을 할 수 있도록 알고리즘이 들어 있기 때문이다.

자동차 불가능했던 산업 현장의 시스템을 인공지능(AI)으로 해결한 국내 연구 사례를 소개한다. 기계 학습사자 사령이 수많은 분류를 하거나, 기계의 고장을 알리기 처리하는 분할할 사자라고 시스템 효율이 올라갔다.

장영재 (반도체, 생산/물류)

✓ 금융/경영

신한금융, 카이스트와 AI 금융 산학협력 체결
 AI 금융 연구센터 설립... 김우창 카이스트 교수 센터장 맡아

신한금융지주는 지난 16일 서울 중구 세종대로에 위치한 본사에서 카이스트(KAIST, 한국과학기술원)와 인공지능(AI) 금융 알고리즘의 공동연구를 위해 산학협력 업무협약(MOU)을 체결하고, '신한-KAIST AI 금융 연구센터'를 설립해 운영한다고 18일 밝혔다.

[제15회 서울국제패션상품컨퍼런스] 김우창 KAIST 산업공학과 교수 "4차 산업혁명 '첨여' 확대" 이끈다

세계-전준범 기자
 입력 2019.04.04 14:52 | 수정 2019.04.04 15:17

카이스트와 국민연금연구원 이목표기반투자(Goal Based Investment) 기술을 활용한 대국민 자산관리 서비스를 개발해 프랑스에서 첫 선을 보았다.

김우창 카이스트 교수와 김정호 경희대 교수, 왕종현 국민연금연구원 부원장, 핀테크 기업 배리노스 등은 2일(현지 시각) 프랑스 파리에서 열린 글로벌 4개 대학(카이스트·프린스턴대·칭화대·에픽비즈니스스쿨) 핀테크 컨퍼런스에 참가해 국민연금·퇴직연금·개인연금을 포함한 자동 재무설계 모델을 시연했다고 밝혔다.

김우창 카이스트 교수와 김정호 경희대 교수, 왕종현 국민연금연구원 부원장, 핀테크 기업 배리노스 등은 2일(현지 시각) 프랑스 파리에서 열린 글로벌 4개 대학(카이스트·프린스턴대·칭화대·에픽비즈니스스쿨) 핀테크 컨퍼런스에 참가해 국민연금·퇴직연금·개인연금을 포함한 자동 재무설계 모델을 시연했다고 밝혔다.

김우창 (금융인공지능, AI/ML을 위한 최적화)

✓ 의료/바이오

'연금술사' 별명, 이상엽 KAIST 교수 "코로나 치료물 질 찾아낼 것"
 입력 2021-09-02 18:24 수정 2021-09-02 23:42

이상엽 KAIST 특론교수(연구부총장사신의 별명은 연금술사. 그의 연구실에서는 썩는 플라스틱은 물론 바이오인크, 식용 색소, 코로나19 치료 후보물질을 다량으로 생산하고 있다. 이상엽의 유전자를 합성해 인간에게 필요한 화학물질을 합성해내는 시스템 대사공학의 창시자가 이 교수다.

이 교수가 보유한 '최초' 기록도 많다. 지난 5월 뉴턴, 아인슈타인 등 역사적인 과학자들이 거처한 과학 아카데미인 영국왕립학회의 첫 한국인 회원으로 선정됐다. 미국공학한림원과 미국국립과학원 회원이면서 영국왕립학회에도 가입한 외국인 회원은 미국영연방 국적을 제외하고는 이 교수가 세계에서 유일하다. 최근 고(故) 정세영 HOC를 명예회장의 이름을 딴 포니형혁신상을 받으면서 최초의 과학 분야 수상이라는 기록도 세웠다.

KAIST 이상엽 교수, 한국인 최초 영국 왕립학회 회원 선정
 2021-08-07 10:29

서울대 입법대학 교수와 공동 편집... "뉴턴-아인슈타인과도 회원"

이상엽 KAIST 교수, AI 이용해 유전정보 복사 단백질 예측
 A 일배 기자 | 승인 2020.12.30 11:45

이 연구팀과 공동개발, 전사인자 예측 시스템 '딥터넷' 개발



이상엽 KAIST 교수, KAIST 재용, 제천대 총장

이상엽 (대사공학, 생물정보학)

✓ 사회/복지

UN APICIT 자문위원 'KAIST 최문정 교수' 위촉
 A 김재혁 기자 | 승인 2021.07.11 11:30 | 승인 2021.07.12 18:55

개발도상국 고령자 삶의 질 개선 위한 ICT 활용 정책 자문

KAIST(총장 이광형)는 최문정 과학기술정책대학원 교수가 유엔 아시아 태평양 정보통신기술교육센터(UN APICIT)의 자문위원(consultant)으로 위촉했다고 11일 밝혔다.

APICIT는 유엔 아시아 태평양 경제사회위원회(UN ESCAP) 산하 단체로 ESCAP 회원국 62개 국가의 정보격차 해소와 인력양성을 위한 정보통신기술 분야 전문 교육기관 및 디지털 국제협력 중간기구다. 2006년에 설립된 한국 최초의 유엔 사무소 소속 기관이다.

APICIT는 사회경제적 발전목표 달성을 위한 ICT 활용에 있어 인적자원 및 제도적 역량을 강화하기 위해 교육, 연구, 정책 자문을 주된 업무 분야로 하고 있다.

[매경 울드포럼] 70대 유투버가 구독자 130만여 명 돌파... "활동적 노화에 관심"
 한복 기자 | 입력 2021.12.08 16:30:26 수정 2021.12.08 17:27:06

경도인지장애(메모리력 감퇴)를 예방하기 위한 인공지능 기반 맞춤형 인지훈련 프로그램 '기억다방'을 소개하는 마켓팅 행사에 참가하고 있다. (가림) 유투버가 구독자 130만여 명 돌파... "활동적 노화에 관심"

최문정 (인공지능윤리, 데이터기반 복지)

데이터사이언스 방법론 연구역량

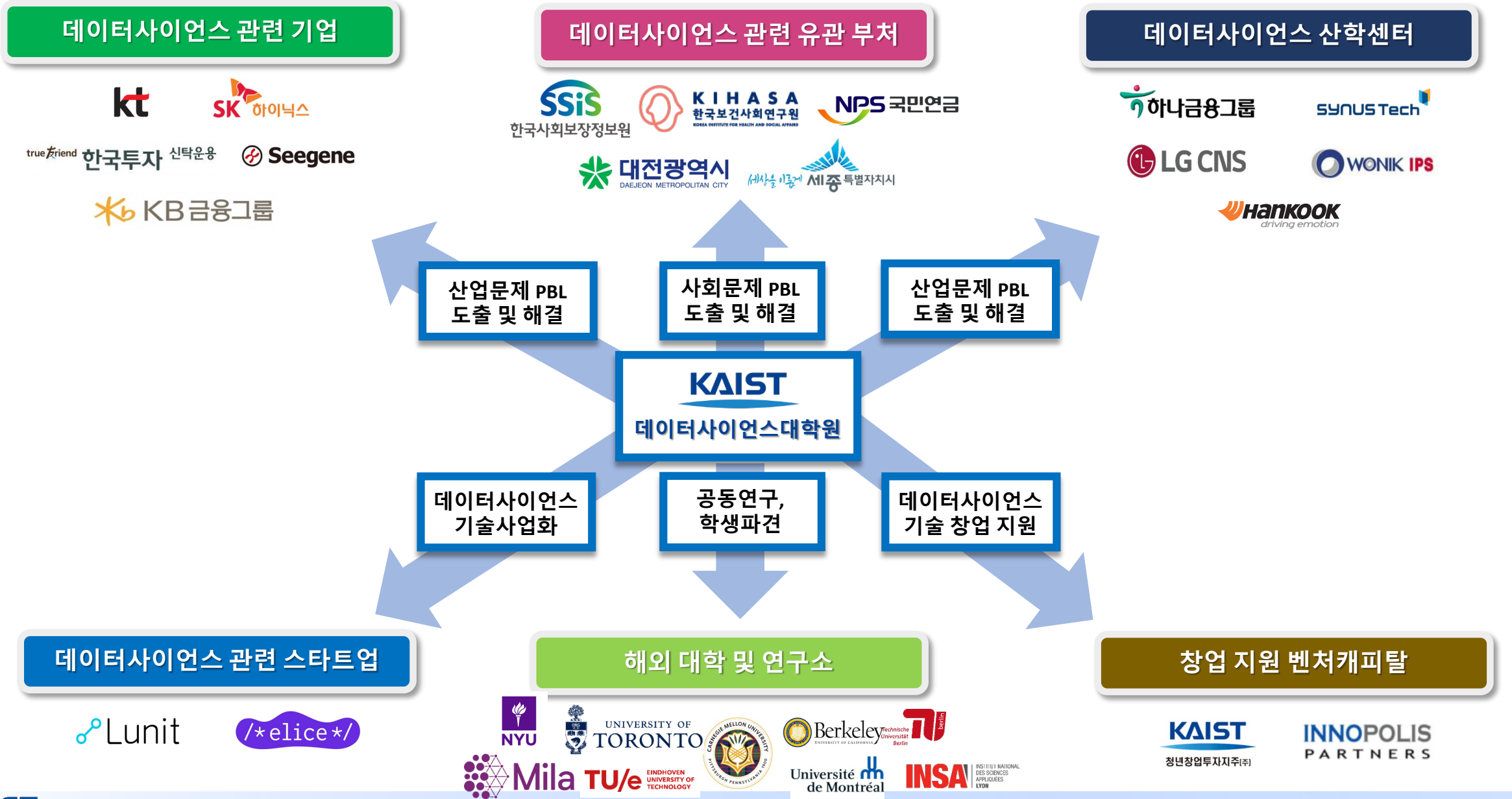
✓ 최근 5년간 우수학술대회 발표 및 학술지 게재 성과

구분	학회 및 학술지 (논문 편수)
데이터 컴퓨팅	SIGMOD (3), VLDB (2), CIKM (4), ICDE (2)
데이터 모델링/분석	KDD (6), ICML (5), NeurIPS (7), ICLR (4), AAI (14), IJCAI (2), SIGIR, WWW (4), IC DM (4), RecSys, WSDM (2)
데이터 활용/가치 창출	CHI (18), Ubicomp, PACM IMMUT (13), CSCW (5), MobiCom, TACL, ACL (3), EMNL P (11), NAACL, UIST (2)

✓ 주요학회 Program Committee 및 학술지 Editorship 활동 내역

구분	활동 내역
데이터 컴퓨팅	SIGMOD, VLDB, CIKM, ICDE PC
데이터 모델링/분석	PAKDD PC Chair, ICDM Area Chair, KDD SPC, AAI SPC, PAKDD SPC, NeurIPS SPC, NeurIPS Tutorial Chair, ICLR Program Chair, NeurIPS/ICML/I CLR/ICDM/WWW PC, IEEE TKDE Associate Editor
데이터 활용/가치 창출	Ubicomp PC Chair, ICMU General Chair, CHI PC, PACM IWMUT Editor, IJ HCS Associate Editor, Transactions of ACL Action Editor, ACL Area Chair, E MNL P Area Chair

프로젝트 협력 기관 및 Coop 협력 기관



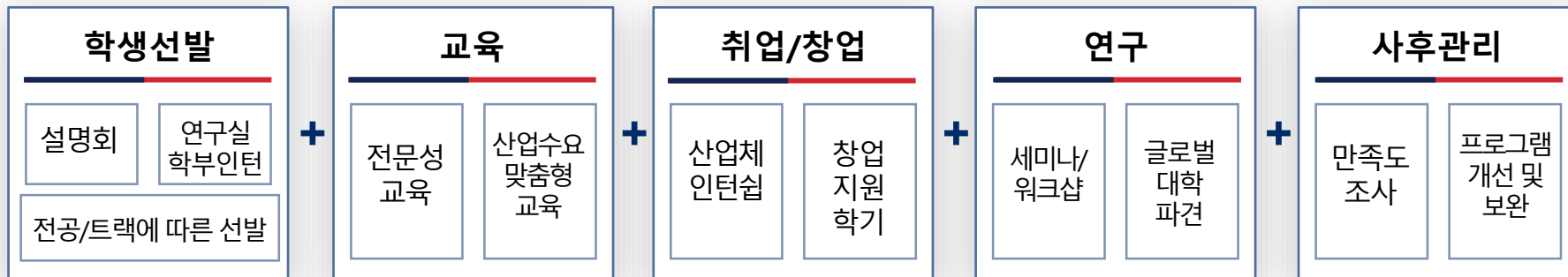
KAIST 데이터사이언스 대학원 이수요건 및 커리큘럼

KAIST 데이터사이언스 인재양성 전략

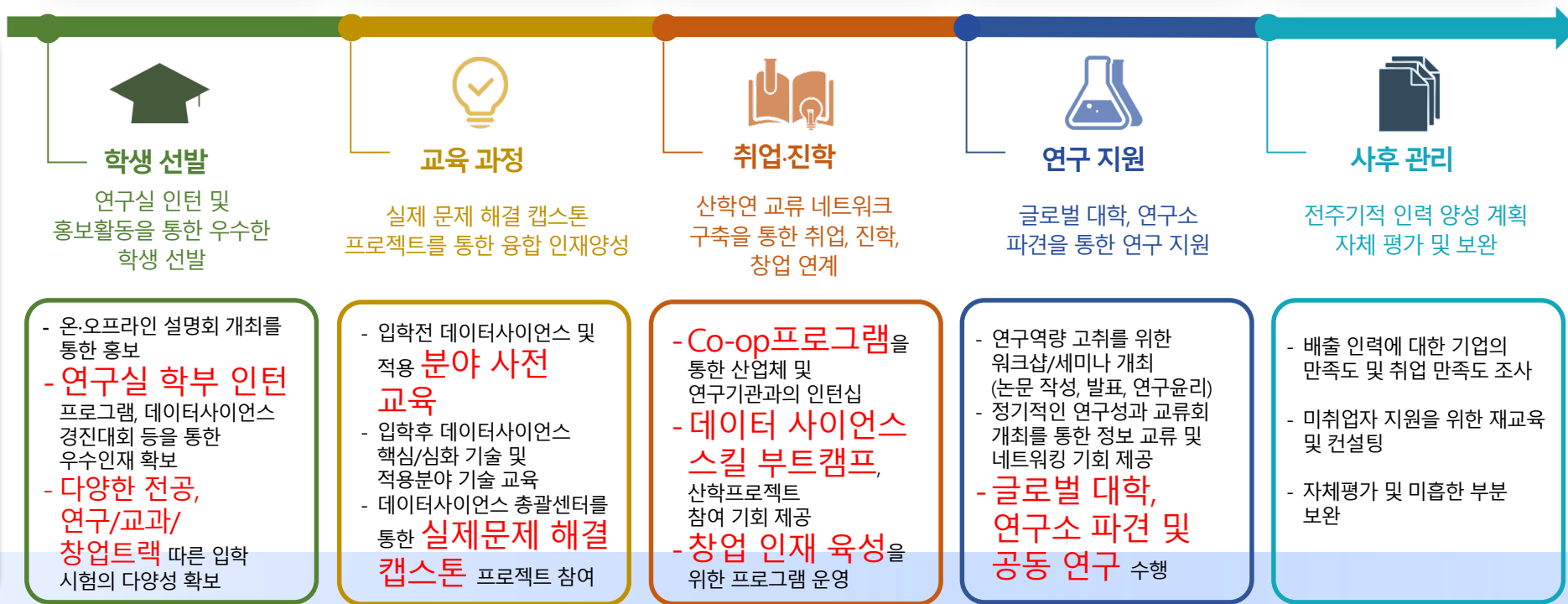
인재양성
목표

산업/사회 수요중심 디지털 혁신 열린인재양성

인재양성
전략



구체적
인재양성
계획



석사과정 대학원생 커리큘럼 이수

✓ 석사과정 이수 요건 기반 시나리오

✓ 33학점

- 9학점 : 연구학점 : 캡스톤 프로젝트 및 디펜스
- 21학점 : 전공선택
 - 15학점 : 데이터사이언스 방법론 과목
 - 6학점 : 데이터사이언스 응용 분야 과목
- 3학점 : KAIST 대학원 공통교과목
 - 리더쉽 강좌, 확률 및 통계, 과학적 글쓰기 등

✓ 부트캠프 활동

- 입학전 부트캠프
- 실무 스킬 부트캠프

✓ 비교과 활동 (기회제공을 의미하며, 필수이수요건은 아님)

- 재학중 산업체/스타트업등 Co-op 기회 제공
- 우수 인재에 대한 해외대학 파견기회 제공
- 창업에 대한 지원 및 투자사 Show Case 제공



- ✓ 데이터사이언스의 활용 및 실무에 익숙한 배경을 가진 학생들을 선발
 - 데이터사이언스 입문 교육 (수리 및 프로그래밍) 을 입학 전에 실시
- ✓ 입학 당시의 평가에 따라 수강의 필요성이 학생 별로 결정
 - 지원자에 대해서도 Pre-커리큘럼의 문호를 개방
- ✓ 실시 시기
 - 9월 입학전인 8월에 "부트캠프"의 형식으로 교육을 제공
 - 9월부터 KAIST 기숙사에 상주하여 교육받을 수 있도록 편의를 제공
- ✓ 과목 형태
 - 입학 사정 과정에서 Pre-커리큘럼의 주제별 교육 필요성을 판정
 - Flipped Learning 형태의 학습 / Hands-on Exercise 형태의 오프라인 실습

부트캠프1.
데이터사이언스
프로그래밍

프로
그래밍

부트캠프2.
데이터사이언스
통계

확률
통계

부트캠프3.
데이터사이언스
수학개론

기초
수학

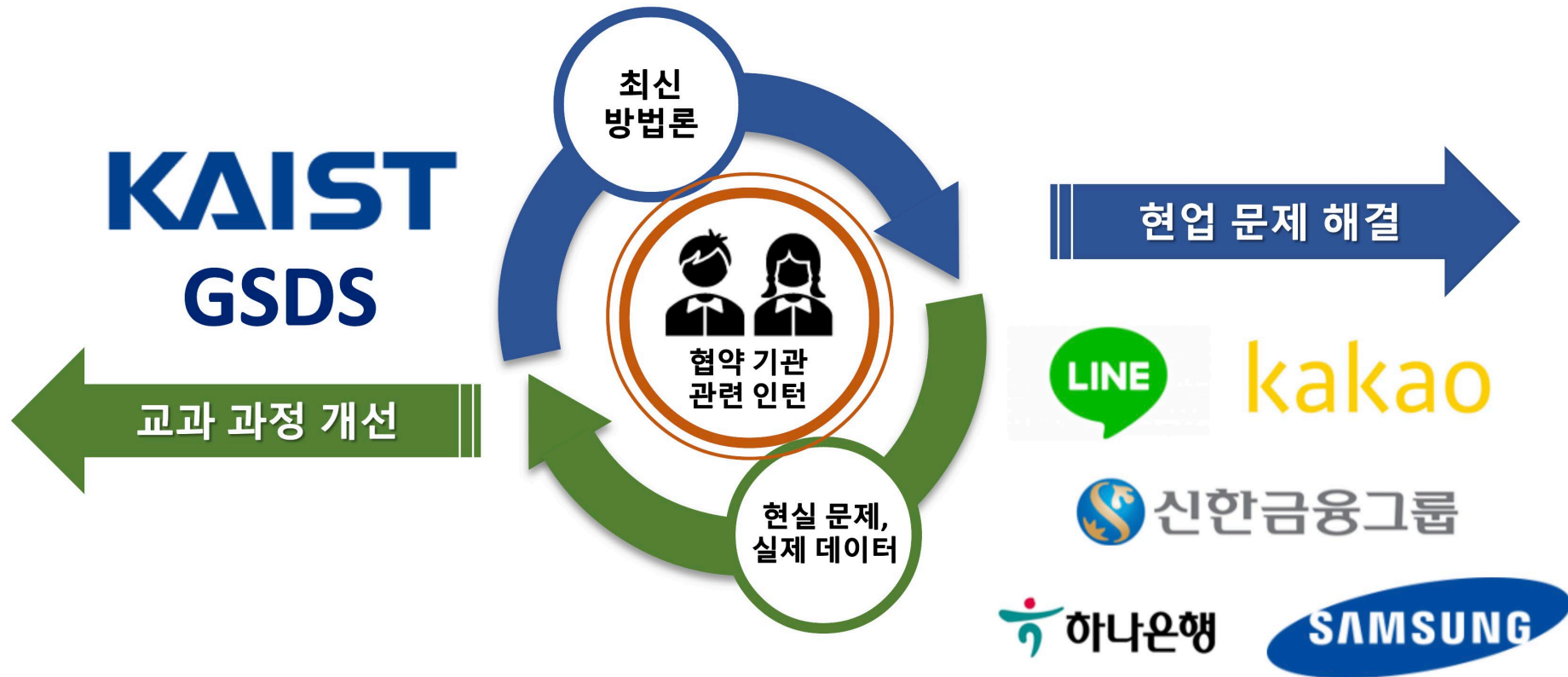
입학전
Pre-
커리큘럼

27개 방법론 과목 (수학, 통계학, 전산, 인공지능 등) +
8개이상의 응용분야과목 (제조, 교통, 의료/바이오, 금융/경영, 사회과학/법학)



✓ 협약 기관 방문 인턴(COOP) 프로그램

- 프로젝트 주제 선정, 지원자 선정, 실제 진행 과정 및 사후 관리 프로세스에 학교가 깊이 관여
- “지속가능한 산학협력 관계 구축” 및 “실제 현장의 문제를 해결할 수 있는 인재 양성”의 목적을 동시에 달성



- ✓ 단계적이며 적극적인 기업체-졸업예정자 매칭 및 중소기업/스타트업으로의 유도



- ✓ 데이터 엔지니어링 부트캠프 (졸업학기 실무교육)

- 산업계 표준 프레임워크 경험 → 대용량 데이터 실무 업무 프로세스 이해
- 방학기간 중 실무자를 초청하여 2주 간의 집중적인 실무교육을 진행

- ✓ 졸업 경력 관리 : GitHub를 통한 PBL 방법론 공개 및 졸업생 평판 제고



- ✓ 데이터로 세상을 바꾸는 유니콘 기업 배출
- ✓ 국내 기술 기반 창업 생태계에 있어서 KAIST가 선도적인 역할을 하고 있음



KAIST창업원에서 주관하는 다양한 창업지원 프로그램 및 행사 내역

✓ 데이터사이언스 대학원 구성원의 적극적인 창업 독려 및 직접적인 지원

- **창업투자사들과의 MOU 체결**
- 아이디어 리그
- 창업 멘토링
- 데이터사이언스 테크데이
- 창업 지원 학기 개설



KAIST
KAIST 청년창업투자지주 X 한국과학기술원
데이터사이언스 인재양성 업무협력 양해각서

카이스트청년창업투자지주와 한국과학기술원(이하 '당 기관'이라 한다)은 산업/사회 수요응답 데이터사이언스 인재양성을 위하여, 인적 교류, 연구 협력 및 우수출원료 다분과 같이 약속한다.

제 1 조 (목적)
이 양해각서는 당 기관의 (가칭) 한국과학기술원 데이터사이언스 대학원 (KAIST Graduate School of Data Science)의 인력양성 및 우수 출원연구에 대한 각종 교류 및 협력 관련 사항을 목적으로 한다.

제 2 조 (운영목적)
1. 이 양해각서에 열거된 사항은 상호 신뢰 및 성실의 원칙에 입각하여 준수한다.
2. 이 양해각서에 열거된 사항으로서 세부 사항이 필요하거나, 열거되지 않은 사항으로 교류 및 협력에 필요한 사항은 당 기관의 제 규정 범위 내에서 별도로 협의하여 정한다.

제 3 조 (교류 및 협력 분야)
당 기관은 제1조의 목적을 달성하기 위하여, 다음 각 호의 사항에 대하여 상호 교류 및 협력한다.
1. 데이터사이언스 인재양성을 위한 인적 교류 방안 도출
2. 데이터사이언스 기반의 청년 창업 투자 확대를 위한 Show Case 및 채워제어
3. 데이터사이언스 기반 청년 창업 투자 연계

제 4 조 (NRF 데이터사이언스 인재양성 지원사업 제안 참여)
당 기관은 제1조의 목적을 달성하기 위하여, 한국과학기술원 데이터사이언스 대학원외 NRF 데이터사이언스 인재양성 지원사업 제안에 협력하사를 보인다.
1) 한국과학기술원에 의해 제안될 예정인 데이터사이언스 인재양성 제안 사업에 선

INNPOPOLIS PARTNERS
이노폴리스파트너스 X 한국과학기술원
데이터사이언스 인재양성 업무협력 양해각서

이노폴리스파트너스(이하 '우'라 한다) 한국과학기술원(이하 '당 기관'이라 한다)은 산업/사회 수요응답 데이터사이언스 인재양성을 위하여, 인적 교류, 연구 협력 및 우수출원료 다분과 같이 약속한다.

제 1 조 (목적)
이 양해각서는 당 기관의 (가칭) 한국과학기술원 데이터사이언스 대학원 (KAIST Graduate School of Data Science)의 인력양성 및 우수 출원연구에 대한 각종 교류 및 협력 관련 사항을 목적으로 한다.

제 2 조 (운영목적)
1. 이 양해각서에 열거된 사항은 상호 신뢰 및 성실의 원칙에 입각하여 준수한다.
2. 이 양해각서에 열거된 사항으로서 세부 사항이 필요하거나, 열거되지 않은 사항으로 교류 및 협력에 필요한 사항은 당 기관의 제 규정 범위 내에서 별도로 협의하여 정한다.

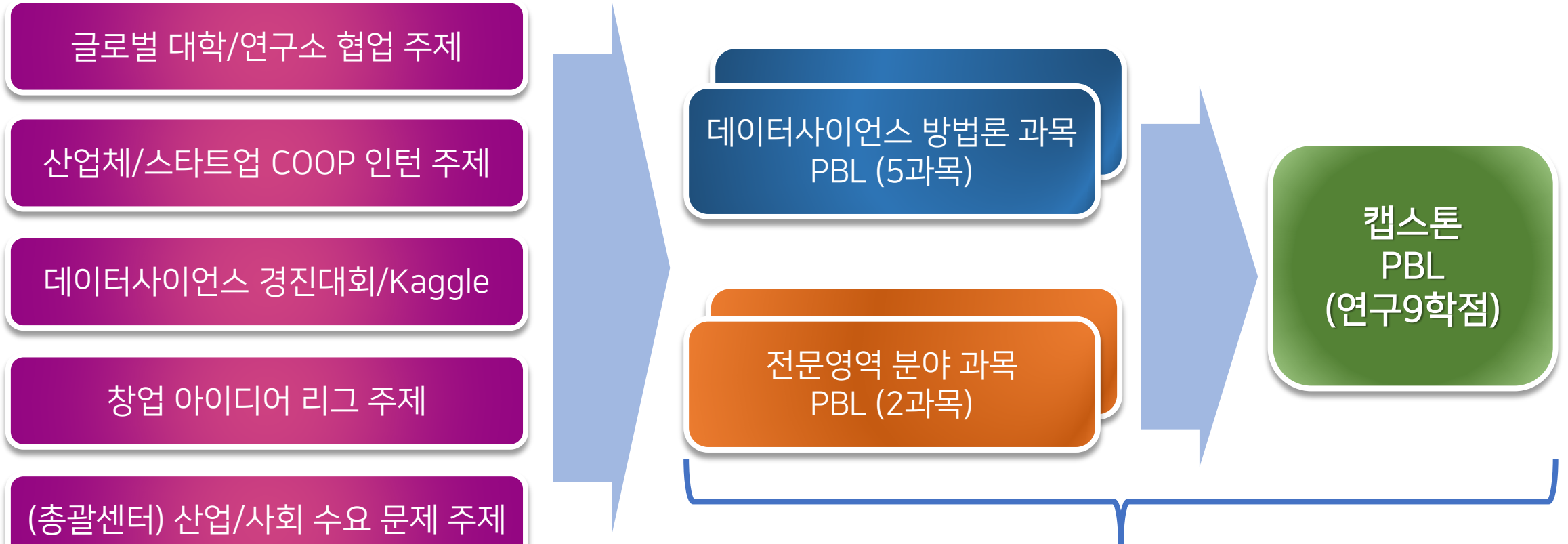
제 3 조 (교류 및 협력 분야)
당 기관은 제1조의 목적을 달성하기 위하여, 다음 각 호의 사항에 대하여 상호 교류 및 협력한다.
1) 데이터사이언스 인재양성을 위한 인적 교류 방안 도출
2) 우수출원 데이터사이언스 연구를 위한 문예 발굴 및 협력 방안 도출
3) 협력기관의 데이터사이언스 관련 교류를 위한 협력 방안 도출

제 4 조 (NRF 데이터사이언스 인재양성 지원사업 제안 참여)
당 기관은 제1조의 목적을 달성하기 위하여, 한국과학기술원 데이터사이언스 대학원외 NRF 데이터사이언스 인재양성 지원사업 제안에 협력하사를 보인다.
1) 한국과학기술원에 의해 제안될 예정인 데이터사이언스 인재양성 제안 사업에 선

창업지원을 위해 창업투자사들과 맺은 업무협력 양해각서 (좌: 카이스트청년창업투자지주, 우: 이노폴리스파트너스)

✓ 현장 문제 발굴 : 산업/사회 수요 문제

- **과목별 PBL** 교육을 통한 접근 : 개별 과목을 통한 본인 주제의 해결 고도화
- **캡스톤 PBL**을 통한 문제 해결 완성 : 캡스톤을 통한 PBL주제의 해결 완료



산업/사회 수요문제 해결을 통한 교육

2024년 가을학기 입시요강

2024년 가을학기 입시 요강

✓ 입학 정원

- 석사 : 00 명
- 박사 : 00 명 (전문연 0명)
- 전원 카이스트 장학생

✓ 원서접수

- 24/3/29 (금) ~ 24/4/11 (목) (17:30 까지)
- <https://apply.kaist.ac.kr/GradApply>

✓ 합격자 발표

- 서류 발표 : 24/05/13 (월) (14:00시 이후)
- 최종 발표 : 24/06/20 (목) (14:00시 이후)

✓ 면접전형

- 24/05/16 – 24/05/26 (서류 합격자에 한해, 날짜 공지 예정)
- 대전 대면면접 (해외 지원자의 경우 비대면 가능)

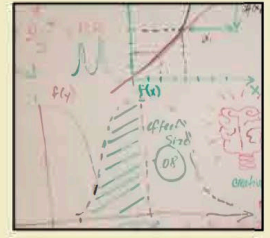
GSDS (데이터사이언스) 대학원 - 석사/석박통합 과정

대기실

면접 준비실
(사전 문제풀이, 60분)

면접실 1
(문제기반 면접, 15분)

면접실 2
(발표기반 면접, 15분)



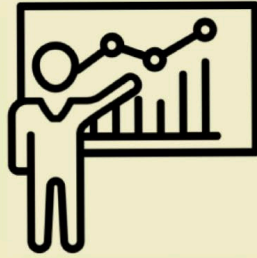
- 통계 및 프로그래밍에 대한 문제풀이 (예시 문제 참고)
- 통계 및 프로그래밍 문제에 대한 구술 면접
 - * 사고 과정과 풀이 결과에 대한 평가
 - * 문제 풀이 시 작성한 답안 메모를 보며 면접 진행
- 지원자 관심분야의 문제 정의 및 소통 능력 평가
 - * 그 분야에 대하여 본인이 생각해 본 문제 발표 (5분)
 - * 본인의 발표 file 사용 가능
 - * Q&A (10분)
- 두 면접실 중 질한 쪽을 기준으로 선발 (두 면접실 점수의 max)
 - * 통계/프로그래밍외의 관심 분야를 가진 학생도 불리하지 않도록 함

면접점수 =
 $\max\{\text{면접실1}, \text{면접실2}\}$

GSDS - 박사 과정 및 재학생석박통합과정

대기실

면접실
(발표기반 면접, 15분)



- 지원자 관심분야의 **문제 정의** 및 **소통 능력** 평가
 - * 그 분야에 대하여 본인이 생각해 본 문제 발표 (5분)
 - * 본인의 발표 file 사용 가능
 - * Q&A (10분)

통계/프로그래밍 문제 sample

Smart기기를 이용하여 숙면에 빠지기까지 걸리는 시간 (확률변수 X)을 조사하여 다음과 같은 data를 얻었다.

Table 1: Time to Sleep

i	1	2	3	4	5	6	...	10,000(= n)	\bar{x}_n
x_i (단위:분)	7.5	3.1	20	5.5	60.0	1.2	...	16.0	10.0

아래의 질문에 답하십시오.

- 데이터 $D = [x_1, x_2, \dots, x_n]$ 을 input으로 받아서, 최소값과 최대값 사이를 k 개로 나누는 bin에 포함되는 관측치의 빈도수를 count하는 histogram을 계산하는 함수를 완성하십시오.

HISTOGRAM (D, k)

Input D : array, $D[i] = x_{i+1}, i = 0, \dots, n - 1$

k : # of bins, $k > 0$.

Output H : array $H[i]$: count for bin $i, i = 0, 1, \dots, k - 1$

- 확률변수 X 의 확률밀도함수를 $f(x) = \lambda e^{-\lambda x}$ 라고 할 때, 위의 데이터로 부터 λ 를 점추정(point estimation)하고자 한다.

(a) 최우추정법(MLE)에 의한 추정량 $\hat{\lambda}_n$ 을 유도하십시오.

(b) $\hat{\lambda}_n$ 과 다른 추정량을 하나 제시하고, 어느 추정량이 좋은지와 그 판단의 근거를 말해보시오.

- 앞서 MLE로 구한 $\hat{\lambda}_n$ 에 대한 다음의 질문에 답하십시오.

(a) $E[\hat{\lambda}_n]$ 를 구하십시오.

(b) $\hat{\lambda}_n$ 는 biased되어 있는가?

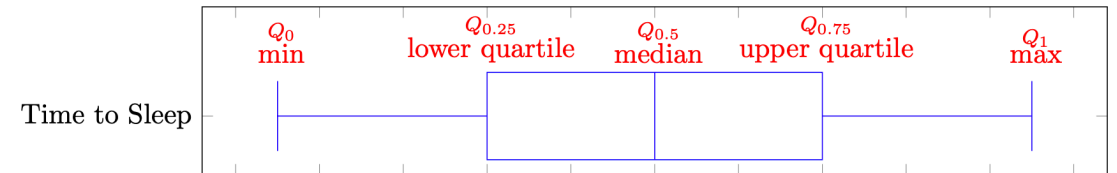
(만일 $E[\hat{\lambda}_n]$ 를 구하지 못했다면, Jensen's Inequality를 이용하여 bias여부를 확인해도 됨)

(c) $\hat{\lambda}_n$ 는 consistent한가?

(d) $\hat{\lambda}_n$ 이 biased되어 있다면, bias를 보정한 불편추정량(unbiased estimator)을 제시하십시오.

- 이 기기는 battery절약을 위해서 60분이 지나면 $x_i = 60$ 으로 기록하고 숙면 monitoring을 멈춘다. 이를 고려하여 λ 에 대한 추정량을 MLE (Maximum Likelihood Estimation)로 유도하십시오.

- 위의 data를 box plot으로 나타내고자 한다.



(a) X 의 분포가 앞서 가정한 지수분포가 맞다면, median과 mean중 어느 것이 더 클까?

(b) X 의 분포가 앞서 가정한 지수분포가 맞다면, $Q_{0.25} - Q_0$ 과 $Q_1 - Q_{0.75}$ 중 어느 것이 더 클까?

(c) 데이터 $D = [x_1, x_2, \dots, x_n]$ 으로 부터 box plot을 계산하기 위한 함수를 구현하기 위한 방법을 제시하십시오. (아래의 예시 답안과 다른 효율적인 방법 제시 요망.)

```
def box_plot(D):
```

```
    N = len(D)
```

```
    sort D by ascending order
```

```
    Q0=D[N*0]
```

```
    Q0.25=D[round(N*0.25)]
```

```
    ...
```

```
    Q1=D[N-1]
```

```
    return Q0, Q0.25, Q0.5, Q0.75, Q1
```

GSDS 면접 준비를 위한 참고 도서

✓ Data Structure Textbooks

- 예1) Tamassia et al., Data Structures and Algorithms in Python

✓ Statistics Textbooks

- 예1) Wackerly et al., Mathematical Statistics with Applications
- 예2) Hogg & Craig, Introduction to Mathematical Statistics
- 예3) 배도선 외, 통계학: 이론 및 응용

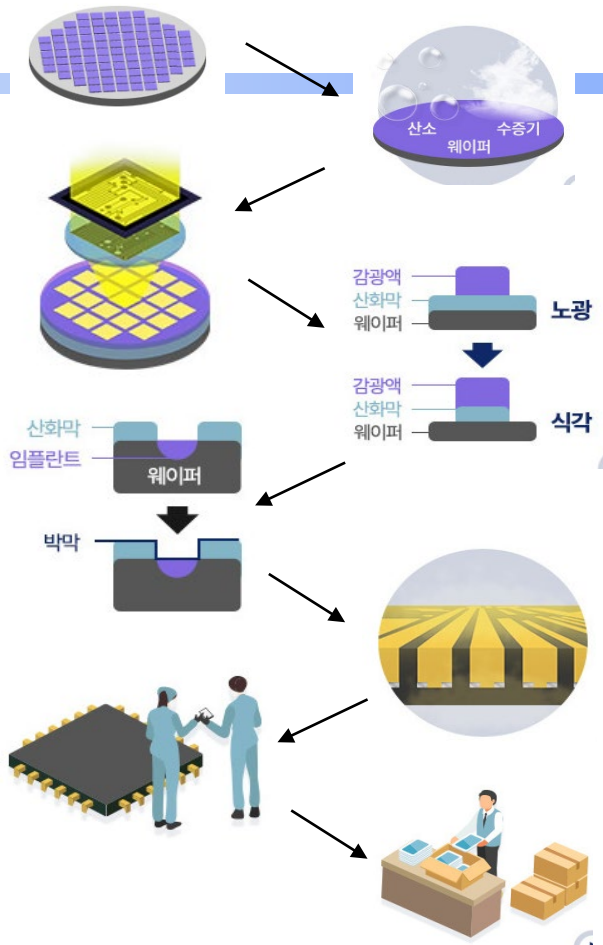
데이터 기반의 반도체 장비 건전성 평가

지원자 : 홍길동

반도체

• 반도체 생산공정

- 웨이퍼 제조
- 산화
- 포토
- 식각
- 증착
- 금속 배선
- 테스트
- 패키징



• 반도체 장비

- 각 생산 공정을 담당하며 팹의 생산성, 품질에 많은 영향을 끼침.
- 장비 제조 업체: Applied Materials, Lam Research, ASML, Tokyo Electron, KLA, Wonik IPS 등



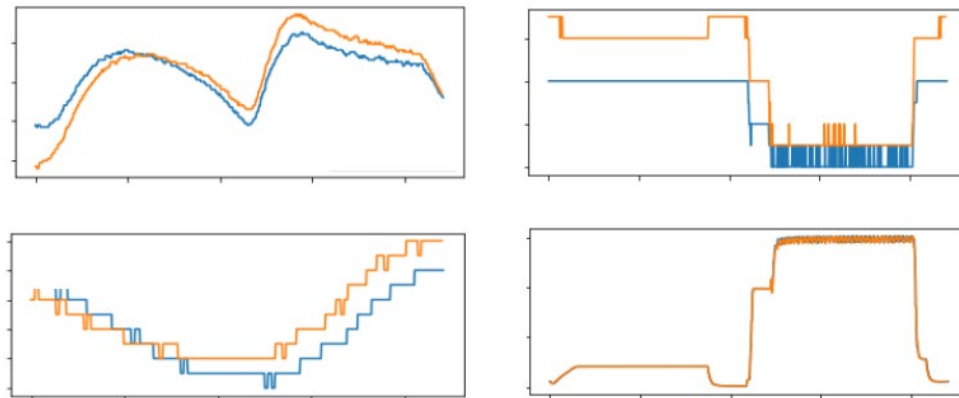
Source: <https://esg.tsmc.com/en/update/greenManufacturing/caseStudy/46/index.html>

<https://www.harting.com/KR/en-gb/node/26158>

<https://news.samsungsemiconductor.com/kr/%EB%B0%98%EB%8F%84%EC%B2%B4-%EB%B0%B1%EA%B3%BC%EC%82%AC%EC%A0%84-%EB%B0%98%EB%8F%84%EC%B2%B4-8%EB%8C%80-%EA%B3%B5%EC%A0%95-%ED%95%9C-%EB%88%88%EC%97%90-%EB%B3%B4%EA%B8%B0/>

반도체 장비 건전성

- 장비의 데이터로부터 장비가 얼마나 건강한지를 파악하고자 함.
- 문제 정의
 - 장비의 건전성에 관한 지표(Label)를 가진 데이터가 많지 않은 상황에서 활용할 수 있는 방법론 개발
 - 반도체 장비에서 발생하는 다각도 데이터(multi-view data)를 활용한 clustering 기반의 계층적 구조를 가지는 건전성 추정 알고리즘 개발
 - 다각도 데이터
 - 하나의 웨이퍼 공정동안 생성되는 process data
 - 장비 maintenance 사이의 trend data
 - 센서 위치 정보기반의 spatial data
- 수집 가능 데이터 종류
 - 온도, 압력, 전류, 전압, ...
- 문제 특징
 - 건전성을 나타내는 지표에 대한 명확한 정의가 필요함.
 - Label이 있는 데이터가 거의 없음.
 - 실 데이터의 경우 결측치나 이상치가 많음.
- 연구 활용방안
 - 데이터로부터 설비 이상을 감지해 고장 전에 Maintenance 진행
 - 일정 시간 주기로 수행하는 Time-based maintenance 대비 효과적
 - 품질에 영향이 있을 것으로 판단되면 해당 제품 계측을 통한 확인
 - 장비의 상태가 좋지 못할 경우 원인 분석



데이터 예시

기존 연구 분석

- 반도체 장비 건전성 관련 연구
 - Real-time diagnosis of semiconductor manufacturing equipment using a hybrid neural network expert system, IEEE Transactions on Components, Packaging, and Manufacturing Technology-Part C, 1997
 - 핵심 아이디어...
 - Hierarchical indices to detect equipment condition changes with high dimensional data for semiconductor manufacturing, Journal of Intelligent Manufacturing, 2014
 - Predictive maintenance in semiconductor manufacturing, ASMC 2015
 - 핵심 아이디어...
 - Data-driven framework for tool health monitoring and maintenance strategy for smart manufacturing, IEEE Transactions on Semiconductor Manufacturing, 2020
 - ...
- 기존 연구 특징 & 한계점
 - 일반적으로 장비의 상태가 linear하게 감소한다고 가정
 - 다각도 데이터를 활용해서 원익 분석을 한 경우는 없음.
 - ...
 - ...

연구 주제

- Research Question 1
 - 다각도의 데이터를 함께 활용하는 것이 그렇지 않은 경우에 비해 어떤 장점을 가질까?
- Research Question 2
 - 다각도의 데이터를 어떻게 활용해야 최종 계층적 구조의 건전성 추정 알고리즘을 개발할 수 있을까?
- Research Question 3
 - 계층적 구조에서 장비 고장의 원인을 파악할 수 있는 방법은 무엇인가?
- Research Question 4
 - 도메인 지식을 가진 엔지니어와 함께 연구를 할 경우 장비 건전성 평가에 도메인 지식을 어떻게 활용할 수 있을까?
- Research Question 5
 - 개발된 방법론의 검증은 어떻게 할 수 있을까?

관심 분야 및 관련 활동

- 지원자의 연구 관심 분야
 - 반도체 공정 및 설비 관리
 - Manufacturing system의 control에 AI/ML 적용
 - Discrete event simulation과 결합
 - ...
- 관심 분야와 관련된 활동 소개 (본인의 우수성 소개)
 - A 경진대회 참가 (B 학회 주관) – 금상 수상 (2023년)
 - AI/ML 공부 blog
 - Discrete event simulation 개발 : github
 - ...

데이터 기반의 Cold-start 상품추천 시스템

지원자 : 홍길동

상품 추천시스템의 cold-start문제

추천시스템 기법

• 추천시스템기법의 종류

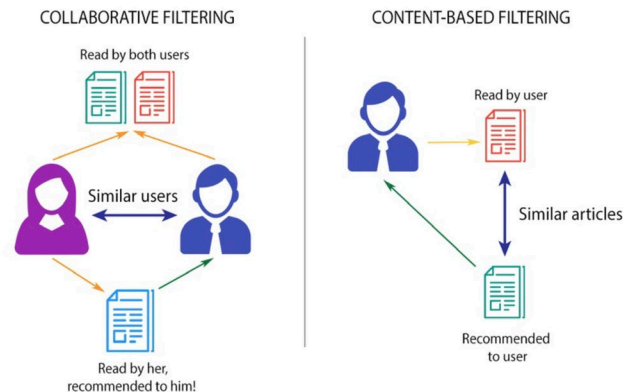
- 컨텐츠 기반 필터링
- 협업 필터링
 - 메모리 기반 방법
 - 모델 기반 방법
- 하이브리드 방법

• 추천시스템에 활용되는 데이터

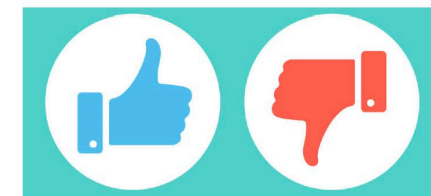
- 명시적 피드백
- 묵시적 피드백
- 외부데이터 (텍스트, 이미지 등)

• Cold-start 상품 추천 시스템의 어려움

- 쇼핑 플랫폼 상에서 신규 상품의 지속적 유입으로 인해 히스토리가 부재한 사용자 및 상품에 대한 정확한 추천이 어려움



명시적 피드백 데이터



Interval-based

Binary

묵시적 피드백 데이터



Click

Add-to-Basket

Purchase

Add song

Cold-start 상품 추천시스템 연구

- 목표: 신규 상품이 지속적으로 유입되는 상황에서 상품의 추천 정확도 향상
- 수집 가능 데이터 종류
 - 상품 클릭/구매/장바구니 ...
 - 상품 설명 텍스트/이미지
- 문제 특징
 - 신규 상품에 대한 사용자의 interaction 데이터가 부재하기 때문에 기존의 협업필터링 방법을 적용하기에 어려움이 존재
 - 따라서, 본 연구에서는 상품의 콘텐츠를 협업필터링 모델에 접목하는 연구를 진행하고자 함
 - 특히, 상품 설명서/이미지 등의 외부 정보를 종합적으로 활용하는것이 목표
- 연구 활용방안
 - 사용자-상품 Interaction 데이터가 부재한 상황에서 제한된 데이터만을 기반으로 정확한 추천을 제공함으로써 소비이력이 많지 않은 상품에 대해 추천을 제공함으로써 사용자의 소비를 유도할 수 있음
 - 이를 통해 신규 사업자들의 적응을 도와주며, 궁극적으로 플랫폼의 수익성을 높일 수 있음

기존연구 분석

- Cold-start 상품 추천 관련 연구

- Wang, Chong, and David M. Blei. "Collaborative topic modeling for recommending scientific articles." KDD 2011
- Ma, Hao, et al. "Sorec: social recommendation using probabilistic matrix factorization." CIKM 2008
- He, Ruining, and Julian McAuley. "VBPR: visual bayesian personalized ranking from implicit feedback." AAI 2016

- 기존 연구 특징 & 한계점

- 기존에 외부 정보를 활용한 cold-start이 존재했지만 이들은 주로 상품에 대한 하나의 modality의 정보를 활용하였으나, 상품에는 다양한 modality 정보가 존재하므로 이들을 잘 결합하여 활용한다면 cold-start 상품 추천시스템의 성능을 높이는 데 큰 도움이 될 것으로 기대함
- ...
- ...

연구 주제

- Research Question 1
 - 상품 설명 텍스트/이미지 데이터를 협업 필터링에 활용하기 위해선 어떤 기법이 필요할까?
- Research Question 2
 - 상품에 대한 다양한 modality의 데이터들의 중요도에는 어떤 차이가 있을까?
- Research Question 3
 - 상품 설명 텍스트/이미지 multi-modal 데이터를 모델링 하기위해선 어떤 기법이 필요할까?
- Research Question 4
 - 개발된 방법론의 검증은 어떻게 할 수 있을까?

관심분야 및 관련 활동

- 지원자의 연구 관심 분야
 - 영화 리뷰 감정 분석
 - 상품 추천 시스템
 - 소셜 네트워크 분석
 - ...
- 관심 분야와 관련된 활동 소개 (본인의 우수성 소개)
 - A 경진대회 참가 (B 학회 주관) – 금상 수상 (2023년)
 - AI/ML 공부 blog
 - 개발 포트폴리오 : github
 - ...

지도교수 배정 방침

✓ GSDS에 참여하는 51분의 모든 교수님께 지도받을 수 있음

- 박사학생은 지원시 기입한 희망지도교수님께 배정하는 것이 원칙
- 석사학생의 지도교수 배정은 happy marriage 방식을 따름
- 모든 참여 교수님께 연간 1명 이상의 TO가 배정되나, 전기에 소진한 분은 후기 신입생 선발이 어려움
- **지원시 희망 지도교수님께 사전 컨택 강력 권장**

✓ 공동지도 적극 권장

- 방법론 + 도메인
- 주지도 교수님 연구실 소속이 원칙
- 필요에 따라 공동지도교수님 연구실에도 소속 가능

✓ 지도교수 결정 시기

- 교수님별 TO는 확정되어 있음
- 입시결과가 나온 직후 지도교수 매칭 시작 예정 (5월말~6월 중순 예상)

FAQ

- ✓ ISysE 소속으로 입학하는 것과 GSDS로 입학하는 것의 차이
- ✓ 사전컨택이 필수인지?
- ✓ 수업시간 관련 (주간.야간)
- ✓ 파트타임 가능 여부 (직장과 병행가능여부)
- ✓ 석박 통합과정 지원자에게 더 요구되는 자질
- ✓ 학점 커트라인
- ✓ 전문연구요원 티오
- ✓ 작년 선발 인원 (석사/석박통합)

경청해주셔서 감사합니다.