



Transform **D**ata Into **B**usiness
Discover New Insight

[DBdiscover.com](https://dbdiscover.com)

2023

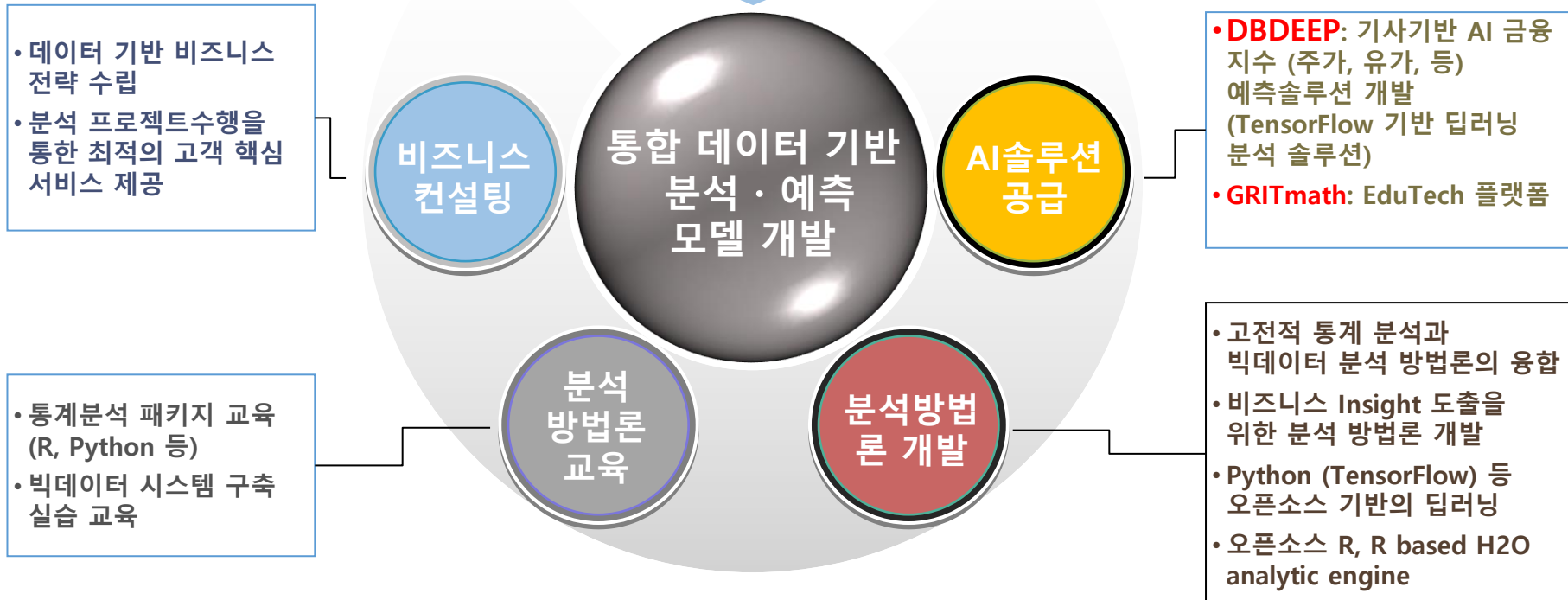
Transform **D**ata into **B**usiness Discover New Insight

DBdiscover는 활용 가능한 모든 데이터를
수집, 분석하여 인사이트를 제공하는
비즈니스 분석 전문기업입니다

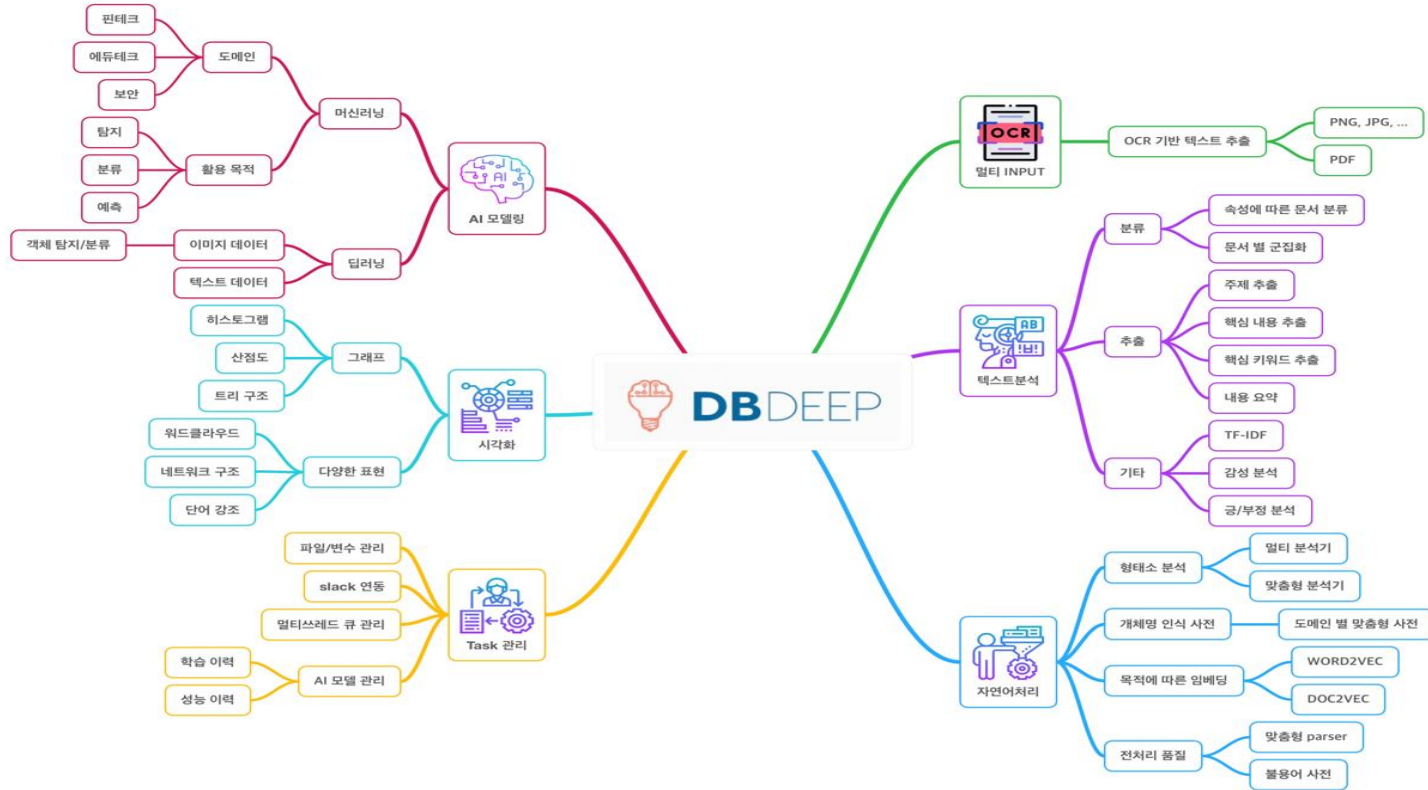
| | |
|------|--|
| 회사명 | 디비디스커버코리아 주식회사 |
| 대표이사 | 김정일 |
| 설립일 | 2013년 2월 25일 |
| 사업분야 | 빅데이터 분석·예측 활용, 인공지능 활용 비즈니스 모델 컨설팅, 학술·연구용역, 소프트웨어 개발 |
| 주소 | 서울특별시 영등포구 국제금융로 106, 행진빌딩 409호 |
| 전화번호 | 02-785-7627 |
| 홈페이지 | www.DBdiscover.com |

DBdiscover (디비디스커버코리아) 는 통합된 데이터를 활용하여 고객이 새로운 비즈니스와 잠재된 역량을 도출할 수 있도록 프로젝트 수행 및 교육 그리고 시스템 구축을 통해 고객의 핵심 가치(Value)를 창출하는 분석 전문 기업

빅 데이터 분석 기반 비즈니스 모델 및 시스템 구축

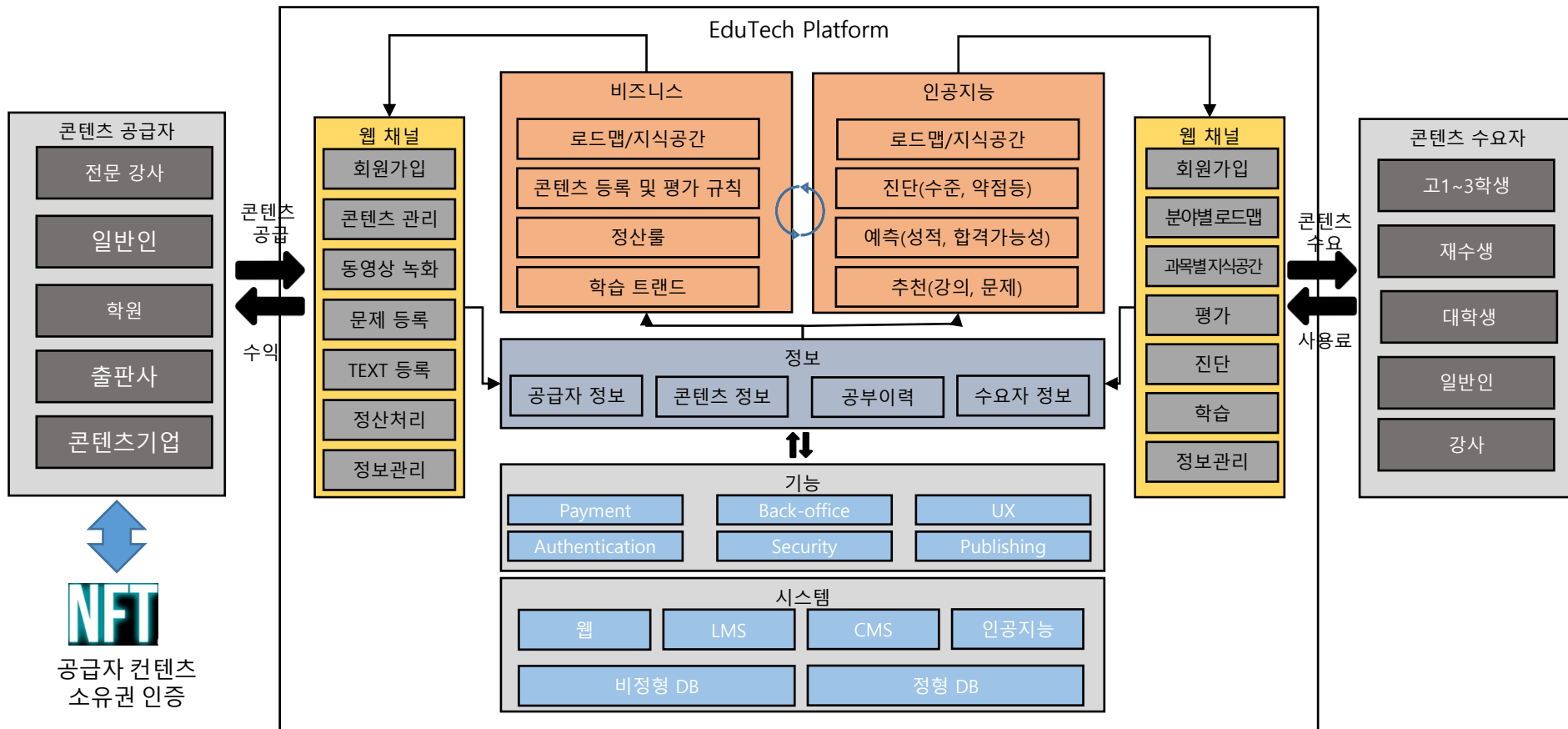


인공지능 기반 텍스트분석을 위한 솔루션 보유: 자연어분석 솔루션 (NLP, OCR, AI모델링)



- 데이터 분석 전문기업 DBDISCOVER의 통합 AI 솔루션 브랜드
- 자연어처리, 문자인식 등 인공지능 기술 기반 고객 맞춤형 AI 솔루션
- 텍스트 데이터에 자체 개발한 인공지능 기반 자연어처리 기술을 이용해 정보 검색 및 추출, 텍스트 분석, 텍스트 요약, 문서 분류를 진행하는 솔루션
- 문자가 포함된 이미지 데이터에 자체 개발한 인공지능 기반 문자 인식 기술을 이용해 문자 인식, 추출 및 편집 가능한 디지털 데이터로 변환해주는 솔루션

- 공급자는 정해진 학습 규칙 및 등록 규격(마이크로 강의)에 맞추어 웹채널을 활용하여 콘텐츠를 등록
- 수요자는 본인이 공부하고 싶은 분야를 선택하여, 공부 로드맵과 과목별 지식 공간에 따라 등록된 콘텐츠와 문제를 풀고 본인의 실력을 증대 시킴
- 수요자는 콘텐츠 사용료를 플랫폼 업자에게 지급하고, 플랫폼 업자는 콘텐츠 사용자에게 사용료를 지급
- 수요자와 공급자의 정보와 공부 이력을 활용하여 AI를 통해 학습자에게 최적의 학습 코스 & 콘텐츠 추천을 진행하여 WinWin.





| | | | | | |
|---|----|------|---|----|---|
| KB금융그룹, 데이터 분석 아카데미 | 02 | 2017 | ○ | | |
| 신한은행, 인공지능 기반 금융지수 예측 모델 개발 | 03 | | | | |
| 현대해상보험, 인공지능 컨설팅 | 04 | | | | |
| ITPP, 딥러닝 기반의 SNS 역기능 개선을 위한 서비스 및 플랫폼 개발 | | | | | |
| SKC, Smart Factory 불량 Loss 관리 | 05 | | | | |
| 교원, 인공지능 수학 상품 컨설팅 | 07 | | | | |
| SK이노베이션, Intelligence Platform 분석(2017) | | | | | |
| 한국산업단지공단, 빅데이터 활용 시범서비스 구축 사업 | 12 | 2018 | ○ | 01 | 남동발전, 빅데이터 플랫폼 구축을 위한 사전분석 |
| | | | | 01 | SK이노베이션, Intelligence Platform 분석(2018) |
| | | | | 02 | 삼성전자, 수요예측 분석 |
| | | | | 05 | 교원, 인공지능 수학 상품 개발 |
| | | | | | |
| 대우조선해양, Ecosystem Infra for Smart Ship 4.0 | 01 | 2019 | ○ | | |
| 한화투자증권, 빅데이터 플랫폼 및 분석체계 구축 | 02 | | | | |
| 연세대학교, 병원 머신러닝을 이용한 치매 행동 심리 증상(BPSD)의 예측 모형 개발 | 07 | | | | |
| 부산은행, 데이터 분석 플랫폼 구축 | 08 | | | | |
| KB국민은행, 인공지능을 활용한 상시 감사 지원 시스템 구축 | 10 | | | | |
| 국가정보자원관리원, 인공지능기반 적응형 보안 시스템 구축 | 11 | | | | |
| 중앙선거관리위원회, 제21대 국회의원선거 정책 선거 연구 | 12 | | | | |

| | | | | | |
|--|----|---|------|----|---|
| | | ○ | 2020 | 02 | 천재교과서, 인공지능 수학 (밀크T AI) 개발 |
| | | | | 08 | 천재교육, AI기반 수능 앱 기획 |
| | | | | 09 | ETRI, 병원 의료기기 행위 모델링을 위한 데이터추출 환경 구축 |
| | | | | 08 | 천재교과서, 수학 CoE (AI기반 추천 모델링) |
| 뉴로비트사이언스, 신경 퇴행성 질환 치료제 후보 물질 도출을 위한 AI 솔루션 개발 | 07 | ○ | 2021 | | |
| NH농협은행, AI 도입을 통한 의심거래 보고시스템 | 07 | | | | |
| KB국민은행, 인공지능을 활용한 상시 감사 지원 시스템 (고도화) | 08 | | | | |
| NIA, 이미지 기반 운행 제한 차량 탐지 모델 개발 | 08 | | | | |
| SBA(서울산업진흥원), AI 활용 사업혁신 지원사업 (AI/빅데이터 기반 세무/회계 플랫폼 고도화) | 10 | | | | |
| | | ○ | 2022 | 06 | NIA, 2022년 학습용 데이터 구축사업 (자전거 도로 주행 데이터) |
| | | | | 07 | 중소벤처기업진흥원, 중소기업 성장경로 예측 알고리즘 개발 |
| 전북은행, AI 기반 준법감시 분석(STR/AML) 모델 개발 | 04 | ○ | 2023 | | |
| 광주은행, AI 기반 준법감시 분석(STR/AML) 모델 개발 | 05 | | | | |
| NIA, 2023년 데이터 활용 기획·검증 사업 | 06 | | | | |
| 한국관광공사, 내비게이션 데이터 활용방안 및 알고리즘 연구 | 11 | | | | |
| 호서대학교, 대화형 AI모델 연구 | 12 | | | | |

“ **DBdiscover**는 데이터 시제 기반의 차별화된 빅데이터 분석 방법론을 보유하고 있습니다. ”



STEP1. 빅데이터 분석의 목적 정의

- 뚜렷한 방향(가설) 설정을 통해 분석의 목적을 명확히 합니다.
- 비즈니스 가치를 고려합니다.
- 차별화된 빅데이터 분석 방법론을 사용합니다.



STEP2. 필요 데이터 확보

- 방향(가설) 검증에 필요한 데이터를 먼저 정의합니다.
- 데이터의 연계를 고려하여 각 소스 별 데이터의 활용 환경을 구축합니다.
- 응용 및 확대 적용이 가능하도록 유연한 기반을 마련합니다.



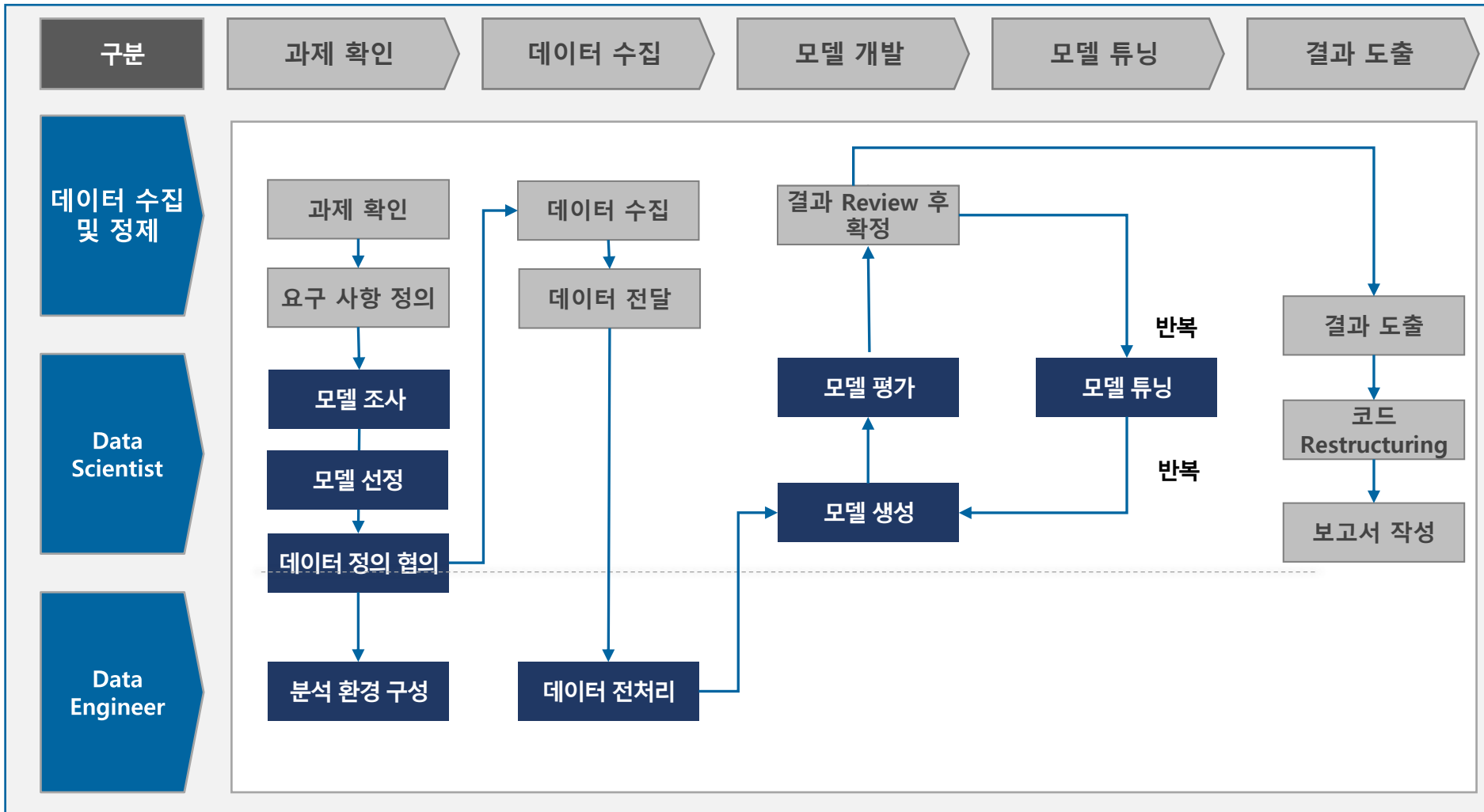
STEP3. 분석을 통한 구현

- 데이터 EDA를 통해 분석 변수의 특성을 파악합니다.
- 데이터의 품질을 반복 검증합니다.
- 데이터 분석 목적에 따라 적합한 방법론으로 분석을 수행합니다.
 - Deep Learning / Machine Learning / Statistical 방법론



STEP4. 가설 및 비즈니스 모델 검증

- 처음 설정한 방향이 올바른지 검증합니다.
- 유의한 요인(factor)을 찾고 영향력을 추정합니다.
- 비즈니스 가치를 분석하고 새로운 비즈니스 모델을 개발합니다.
- 지속 가능한 운영이 가능하도록 운영 프로세스를 정립합니다.



- 특허 : 문항순서 정렬에 의한 개인 맞춤형 콘텐츠를 제공하기 위한 장치 및 이를 위한 방법
- 특허번호 : 제10-2405838호
- 출원일 : 2021.05.03
- 등록일 : 2022.05.31



- 특허 : 인공지능경망 모델을 이용한 뉴스 기사 분석에 의한 주가지수 예측 방법 및 장치
- 특허번호 : 제10-2194200호
- 출원일 : 2018.11.23
- 등록일 : 2020.12.16

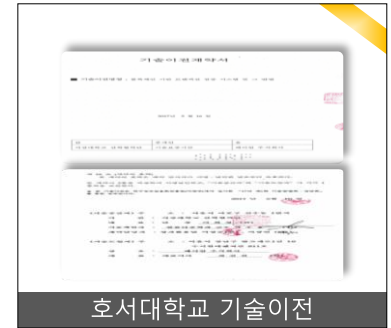
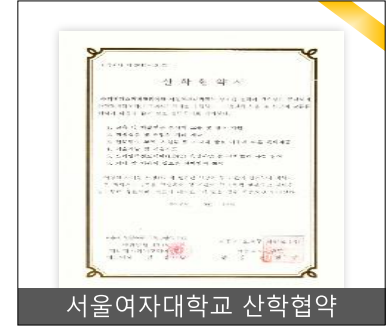
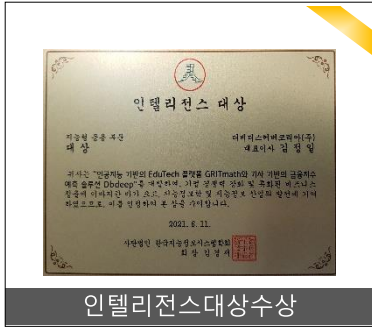


- 특허 : 결정트리 및 심층신경망을 이용한 대량의 데이터를 분류하기 위한 장치, 이를 위한 방법 및 이 방법을 수행하기 위한 프로그램이 기록된 컴퓨터 판독 가능한 기록매체
- 특허번호 : 제10-2184395호
- 출원일 : 2020.05.06
- 등록일 : 2020.11.24



특허 출원 중

| 번호 | 상태 | 특허 |
|-----------------|----|--|
| 10-2021-0010373 | 출원 | 문항 특성 정보가 반영된 기계학습모형 기반 학습 분석 및 개인 맞춤형 콘텐츠 제공 방법 |
| 10-2019-0049637 | 출원 | 학습자 맞춤형 교육 정보 제공 방법 및 장치 |
| 10-2018-0084071 | 출원 | SNS 사용자 간 역기능 개선을 위한 방법 및 장치 |



BUSINESS AREAS



비즈니스 컨설팅

IoT환경과 핀테크를
연결하는 전략 수립
신규 비즈니스 모델
발굴



AI 솔루션 공급

비즈니스 모델 별
맞춤 솔루션 구축

산업별 자체 개발
인공지능 솔루션 공
급



빅데이터 분석/예측

최신 빅데이터 분석 기
법 활용한 데이터 분석
및 예측

비정형데이터와 정형
데이터 융합 분석
전통적 통계 분석방법
론과의 통합적 접근



데이터 분석가 양성 교육

기업체 및 교육기관을
대상으로 빅데이터
분석 교육

맞춤형 커리큘럼을 통
한 빅데이터 기반
비즈니스 분석 교육

AI 서비스 경험

금융, EduTech 및 다양한 도메인의 예측, 탐지, 및 추천 모델링 기반의 AI서비스를 제공함.

데이터 구축, 데이터 분석, 통계/머신러닝/딥러닝 모델 개발, 실시간 서비스 영역 까지 분석에 관련된 전영역에 관한 기술력을 보유.

AI 서비스 구조도

| 대분류 | 금융 | 교육 | 의료 | 제조 | 보안 | 공공 |
|-----|----------|---------------|--------|-----------|--------|--------------|
| 예측 | 고객행동예측 | 수학수준 예측 | 약물식별 | 수요예측 | 전력량 예측 | |
| | 시장예측 | 학습자 행동 예측 | 감염병예측 | 가격예측 | | |
| | 상품지수예측 | 출제 경향 별 문항 예측 | | 장애예측 | | |
| | 판매수요예측 | 성취도 기반 석차 예측 | | | | |
| 탐지 | 이상거래탐지 | | 의료이상탐지 | 공정장애탐지/예측 | 해킹탐지 | 과적차량탐지 |
| | 사기탐지 | | | 제품이상탐지 | 장애탐지 | 불법개조차량탐지 |
| | 이상거래예측 | | | 설비이상탐지 | | 불법심사청구탐지 |
| 추천 | 종목/상품 추천 | 문항추천 | 약물추천 | 원유배합추천 | | 산업단지 내 업체 매칭 |
| | | 경로추천 | 의료추천 | 공정최적화 | | |
| | | 학습자수준진단 | | | | |

“감염병 예측 시스템” 구축을 위한 수행 절차 및 방법론에 대한 일반적 방식은 다음과 같습니다.



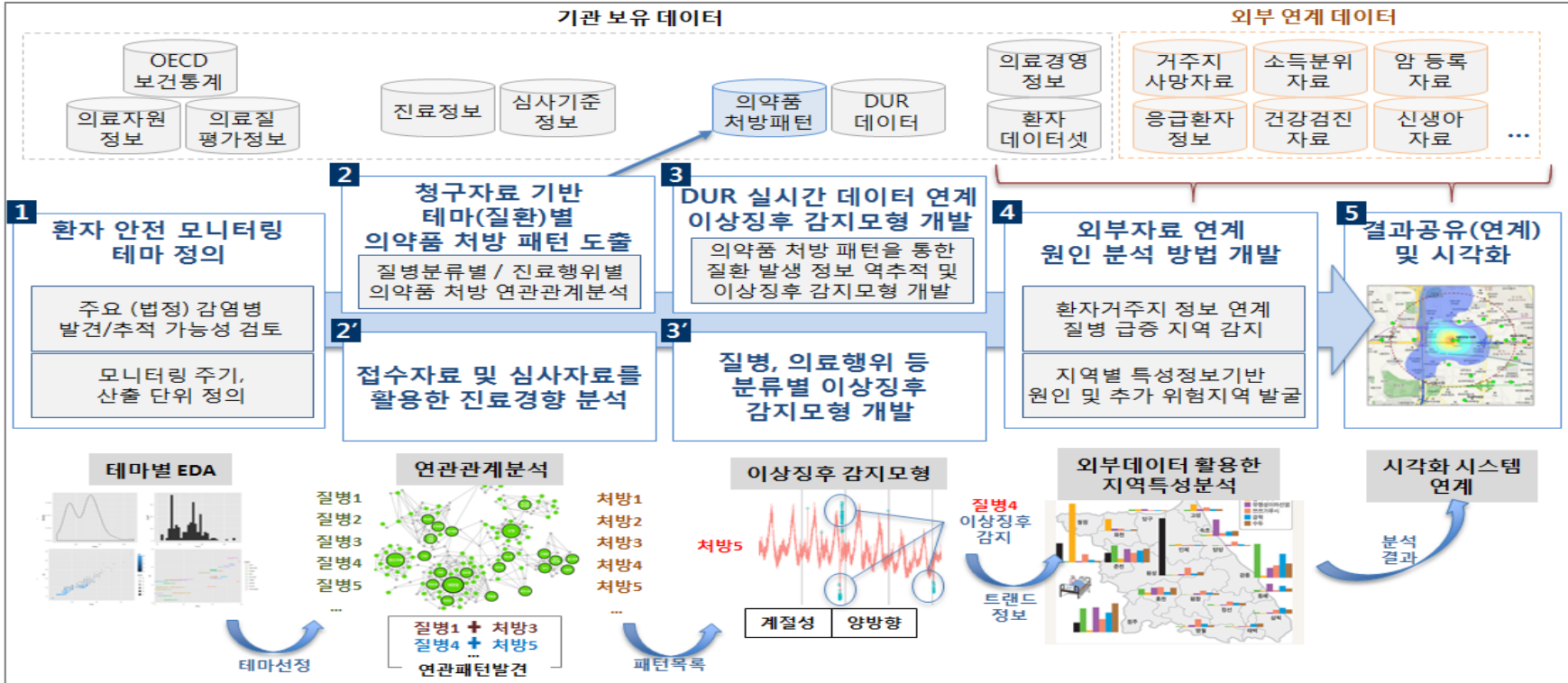
업무 파악 및 테마 정의
 환자안전 모니터링 테마 정의 및 분석 테마별 EDA(Exploratory Data Analysis, 탐색적 데이터 분석)를 통한 데이터 특성 파악

테마 탐색
 과거 축적된 청구자료 및 접수, 심사자료 기반 분석 테마별 연관관계 분석

모형 개발
 DUR 실시간 데이터 및 의료기관 청구자료 기반 이상징후 감지 모형 개발

모형검증 및 모형확장
 환자거주지 및 의료자원 정보 등 연계 분석을 통한 원인분석 방법 개발

정보 공유 및 활용
 GIS 솔루션을 활용한 시각화 시스템 연계 방안 개발

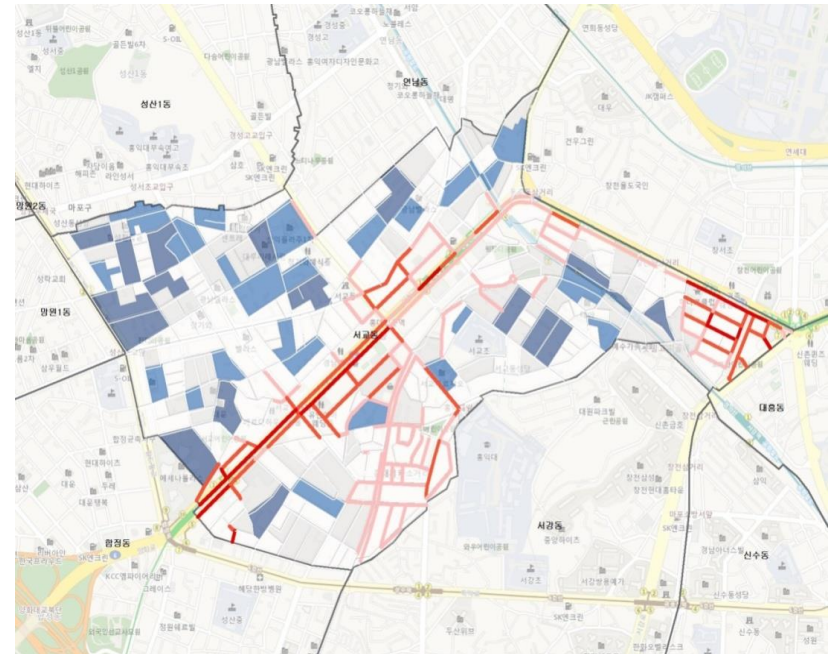
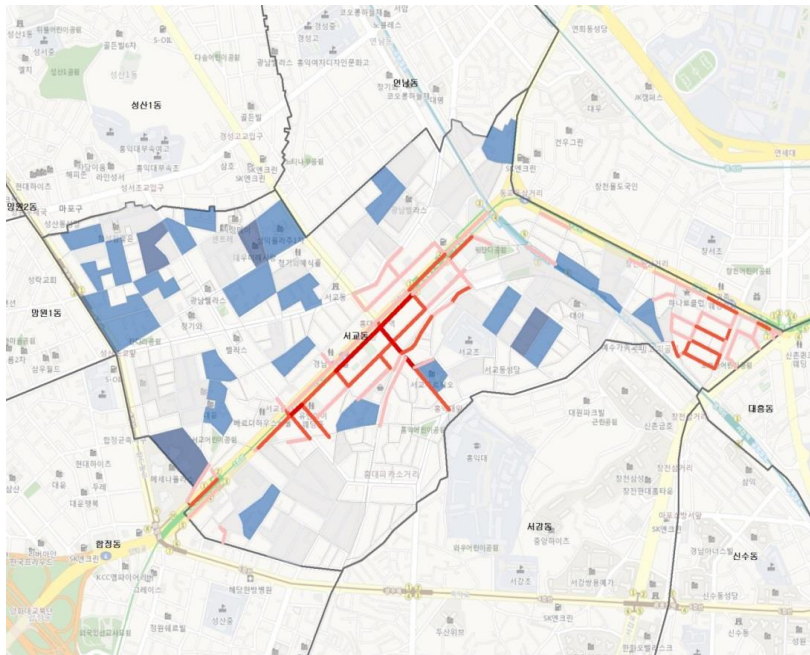


선거데이터 분석:연령대 별 유동인구 및 거주인구 비교, 분석

젊은 유권자 (20대, 30대) 분석

- 서교동의 20대의 유동인구 수는 다른 연령 대비 월등히 높음
- 젊은층이 많이 활동하는 시간대는 오후 6:00 ~ 오후 9:00 임
- 타 연령과 비교 하였을 때 20대와 30대의 거주 지역은 비슷한 분포를 보였으며, 유동이 많은 지역 또한 비슷한 분포를 보였음

20대, 30대를 타겟으로 하는 유세 및 선거 전략은 아래의 지역에서 하면 효과적임

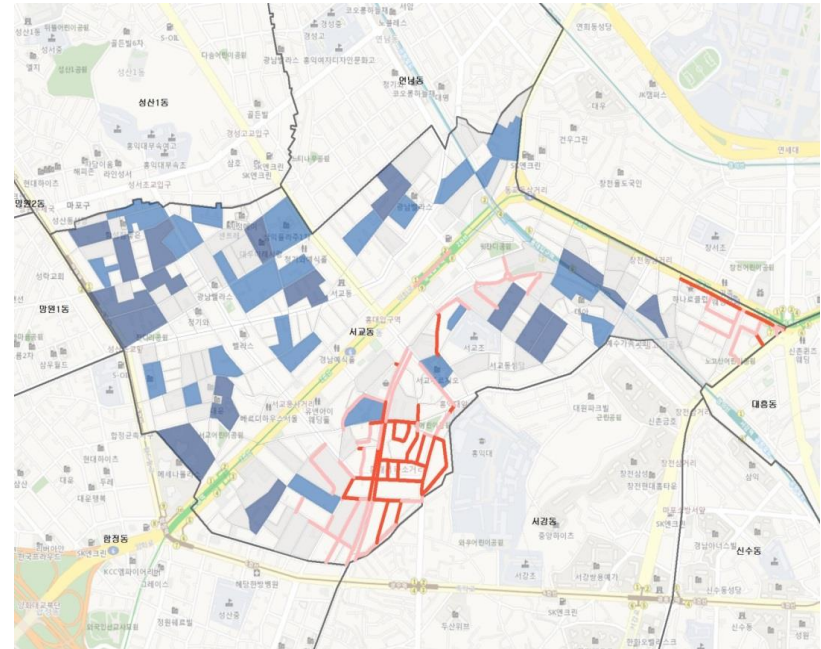
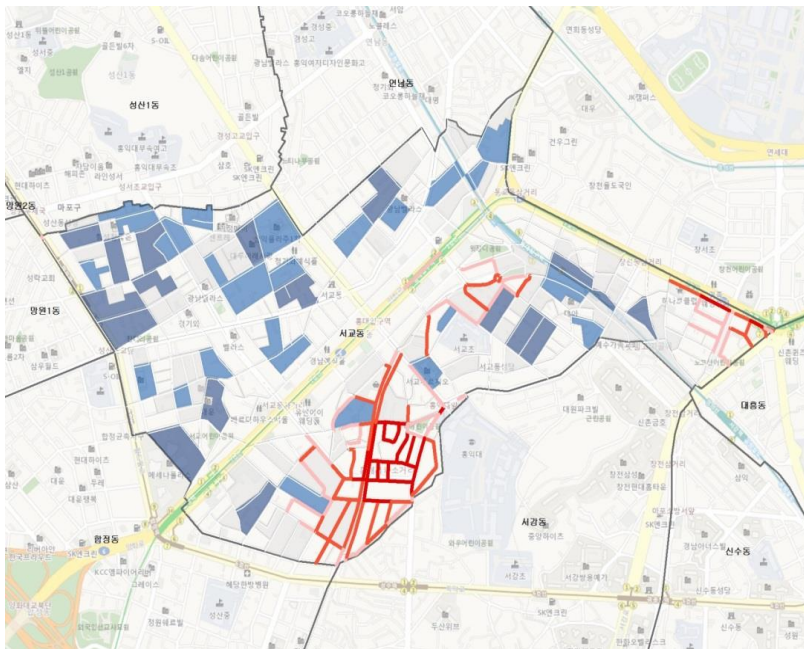


선거데이터 분석:연령대 별 유동인구 및 거주인구 비교, 분석

40대 이상 유권자 분석

- 40대 이상의 유권자들은 오후 2:00 ~ 6:00 에 가장 많은 유동 인구가 있었음
- 젊은층 (20,30대)와 비교하였을때, 거주 지역에 약간의 차이가 있음
- 젊은층과 비교 하였을때 40대이상의 유동이 많은 지역은 확연한 지역 차이를 보임

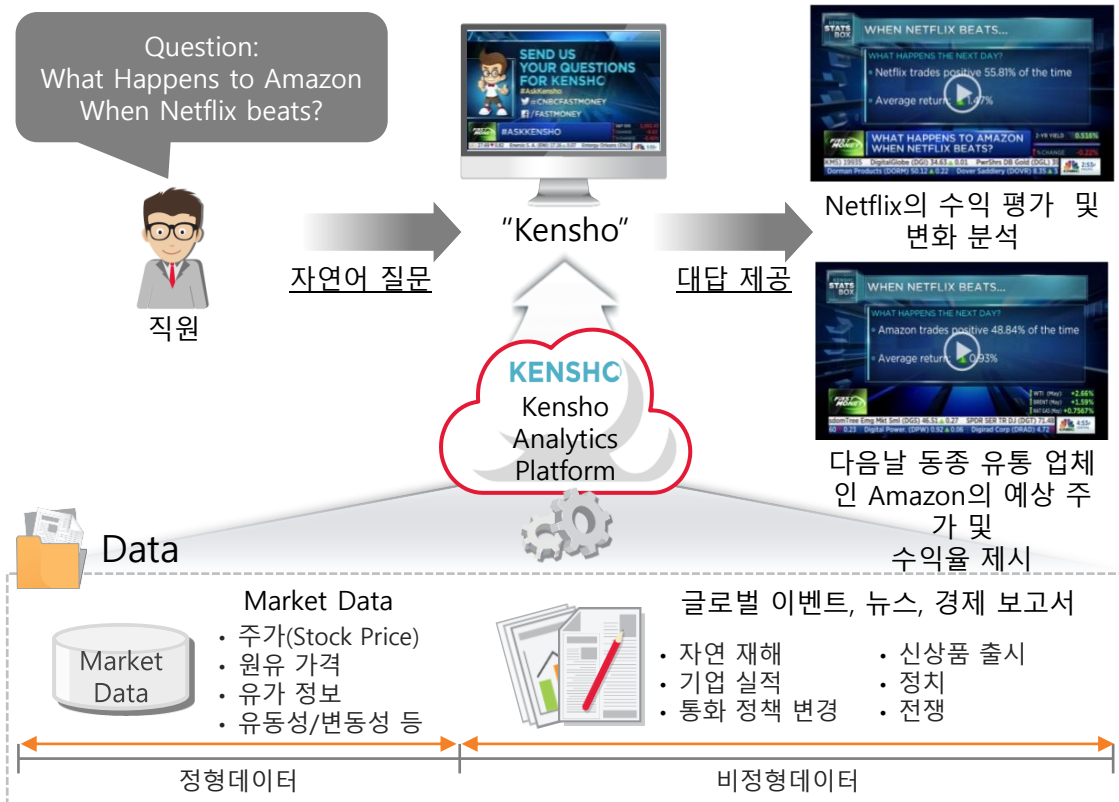
40대 이상의 유권자를 타겟으로 하는 유세 및 선거 전략은 아래의 지역에서 하면 효과 적임



Kensho 사례

- Kensho는 머신 러닝 알고리즘을 활용, Q&A형 대화를 통해 복잡한 금융 관련 질문에 대해 실시간으로 결과를 제공함

CNBC의 "#Ask Kensho" 서비스 예시



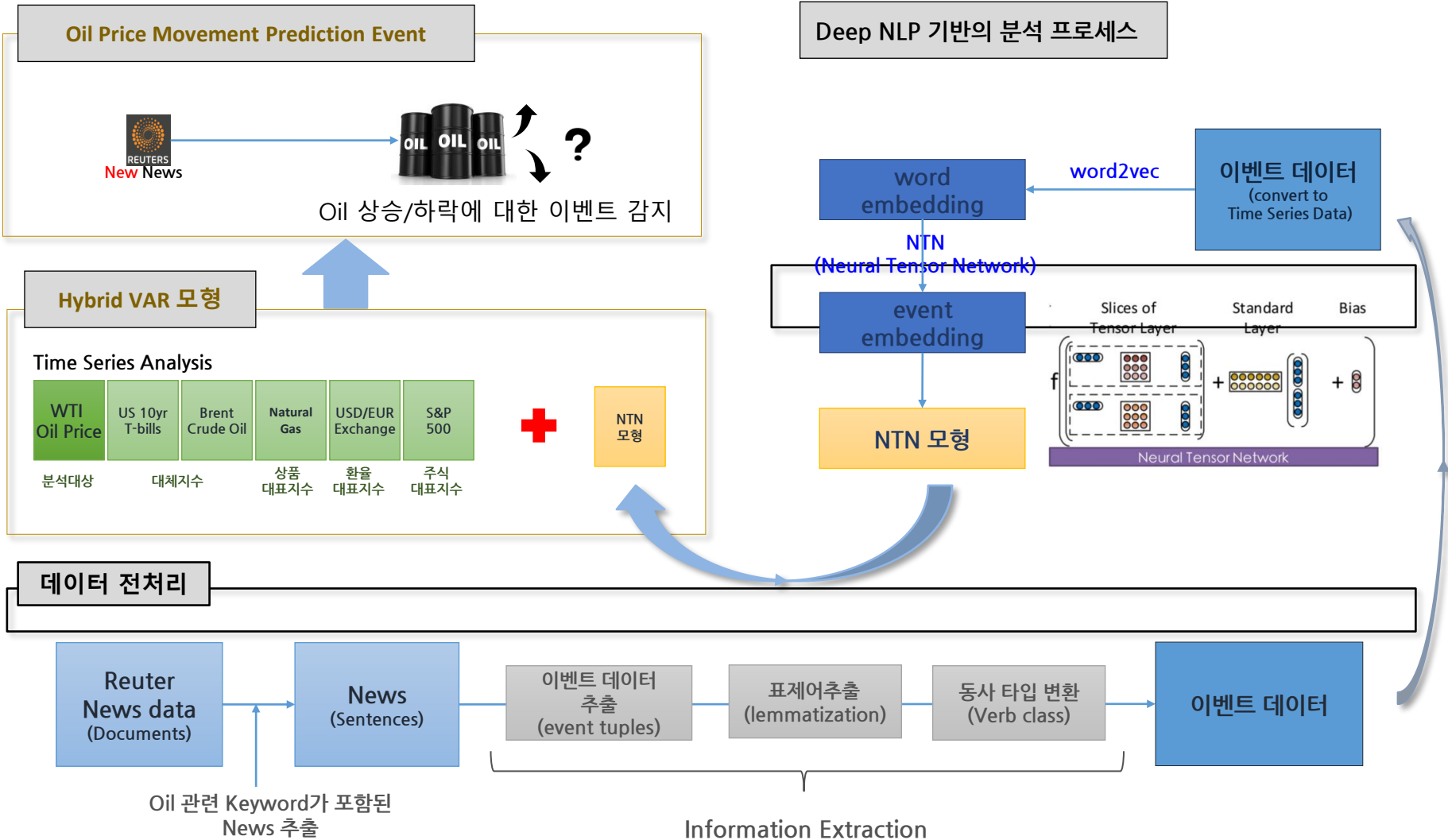
Overview

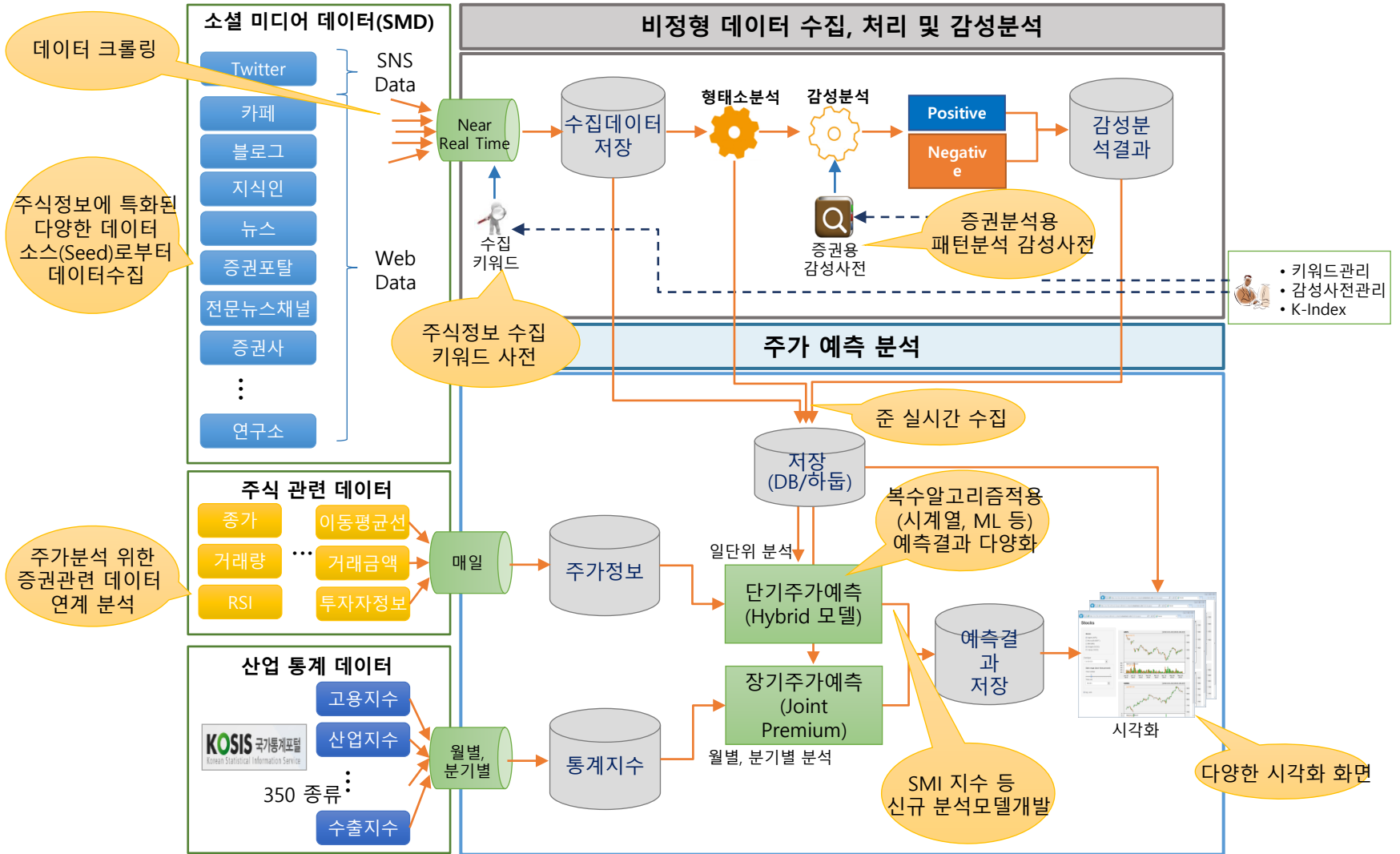
- 본사: Cambridge, Massachusetts
- CoFounder & CEO: Daniel Nadler
- 설립: May 1, 2013
- 직원 수: 50
- 기업가치: \$650M (약 7,200억원)

서비스 현황

- Global Top 금융기업 사용 중
 - Goldman Sachs, JP Morgan 등 다수의 금융 기업에서 사용 중
- 40시간 소요 보고서를 10분내 완료
 - Analyst 리서치/분석 활동 지원 (업무 시간 단축, 정확성 향상)
- 과거 fact 기반의 예측 방식 사용
 - 실제 발생한 History Data를 수집·분석하여 현재에 적용 (9만여개의 Event 정의)

은행 금융시장 지수 예측 수행 2017





은행 인공지능을 활용한 상시감사 지원 시스템 2019

| | | | |
|-----------|---|-----------|--------------------------------------|
| 과제 | 머신 러닝 기반의 가계 대출 사기 예측 모델 머신 러닝 기반의 기업 신용 대출 사기 예측 모델 | 목표 | 기업 모델 AUCPR 0.96 , 가계 모델 AUCPR: 0.83 |
| | | 목적 | 은행 대출 사기 방지 모델 개발 |

분석 대상 대상 테이블 모델링 변수 선정

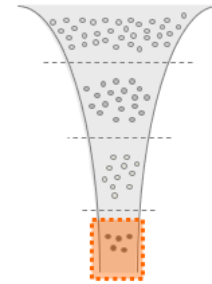
가계 신용 대출 사기 예측

- 직전 3년(2017.01~2019.12)간의 신규 여신 청구 접수 건 중 대출 서류 송부 완료 대출 건
- 사기율은 0.06%로 불균형 데이터

기업 신용 평가 사기 예측

- 직전 8년(2012.01~2019.12)간의 신용평가가 완료된 기업 대출 건
- 사기율은 1.1%로 불균형 데이터

| | |
|-----------|---------------------|
| 인구통계학 | 고객의 연령/성별 등 정보 |
| 재직 및 소득 | 고객의 재직회사 및 소득 정보 |
| 은행 거래 | 당행거래기간, 6개월 평잔 등 정보 |
| 사기대출 속성 | 사기업체명 유사도 등 정보 |
| 대표자 | 기업의 대표자 변경 및 신용 정보 |
| 내/외부 신용평가 | 기업의 재무/비재무 신용평가 정보 |



동일값/결측치/단순정보 등 제거
변수생성(재무정보 퍼넷, 산업평 균차이등)

통계적검증(베르누이분산, 키이 제곱, T-test, 다중공선성처리)

랜덤 시뮬레이션 기법, 베이지 안 최적화 기법 활용

예측 모델 개발 변수 중요도 산출 운영 지원

가계

- 지도학습 기반
- SMOTE 활용 데이터 분균형 처리
- Random Forest/XGboost, Stacking, DNN, CNN 모델 개발
- 최종 XGboost 적용

기업

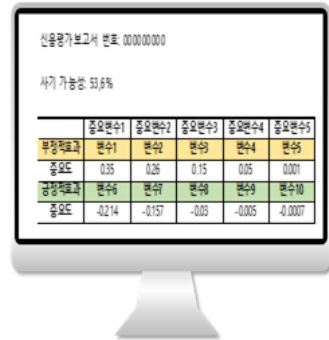
- 비지도학습 기반
- Autoencoder, Isolation Forest, GAN, KCOD 모델 개발
- 최종 Isolation Forest 적용

모델 튜닝

- Bayesian 최적화 방법론 적용



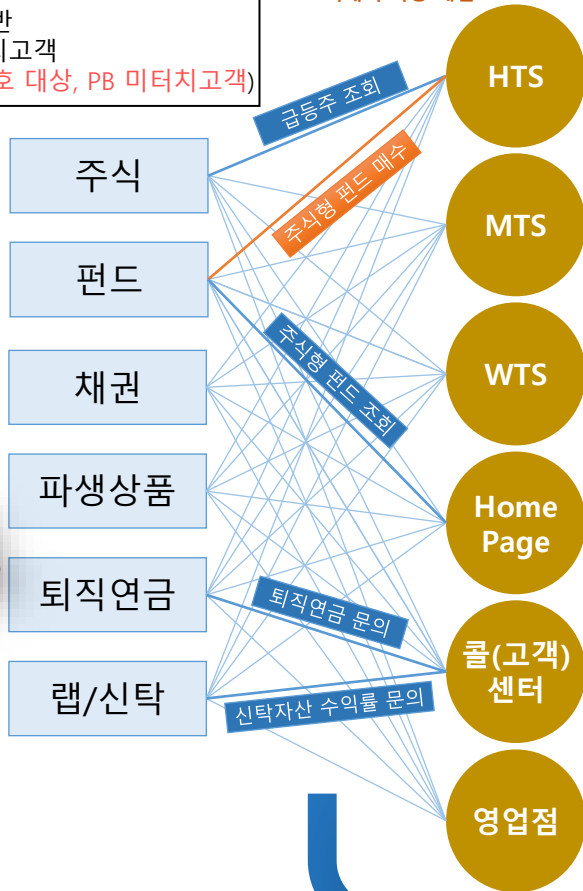
- 연합게임이론의 Shap Value 적용
- 모델 전체의 변수중요도 가이안 개별예측건의 사 기확률및사기가능성에 영향을 미치는 변수들의가 중치 계산



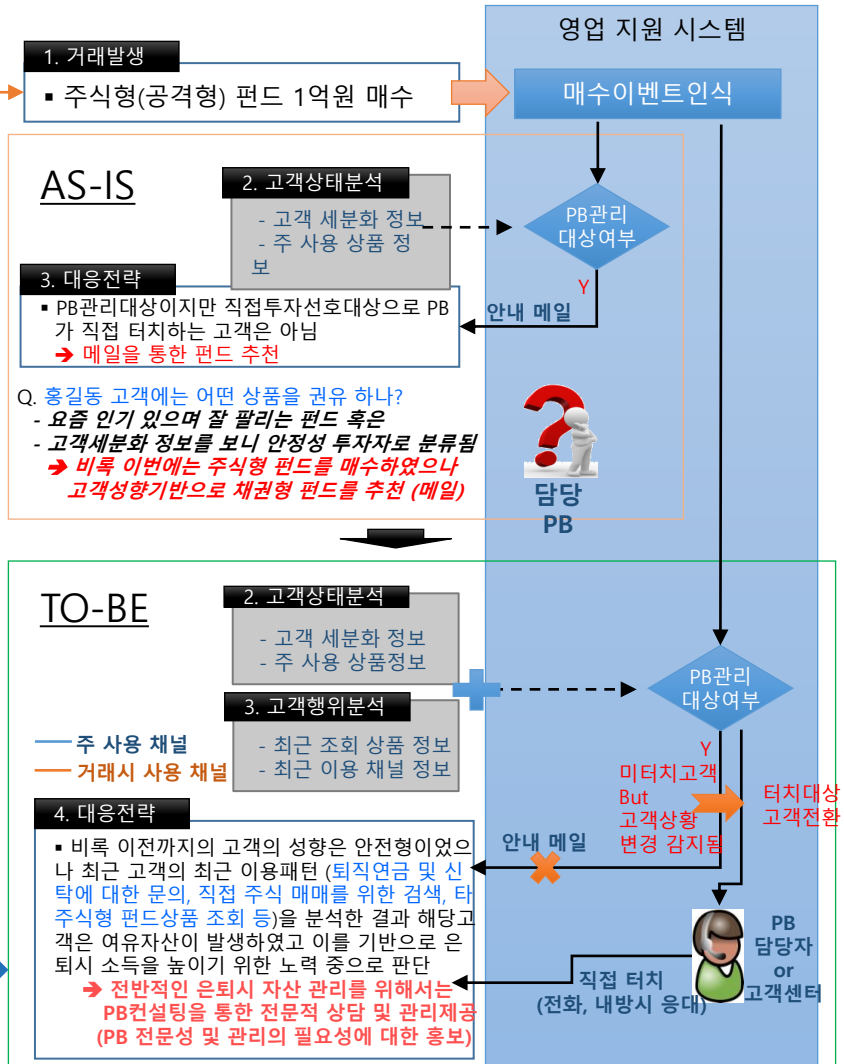
- 개별 대출신용건별예 측결과(사기확률및가 능성에영향을미친 변 수 항목을사전점검화면 에제공하여 사기대출 을미연에방지

김 고객**
 - 연령대 : 50대 중반
 - 고객등급: PB 관리고객
 (But 직접 투자선호 대상, PB 미터치고객)

— 주 사용 채널
 — 거래시 사용 채널



고객 상품 채널이용기록 채널



- AI 수학의 인공지능 부분을 기획, 모델 및 서비스를 개발하여 실시간 맞춤 문항, 맞춤 처방 등의 서비스를 제공

실시간 맞춤 문항

실시간 분석으로 추천 되는 맞춤 문항으로
우리 아이의 수준에 맞는 효율적인 분석

맞춤 처방

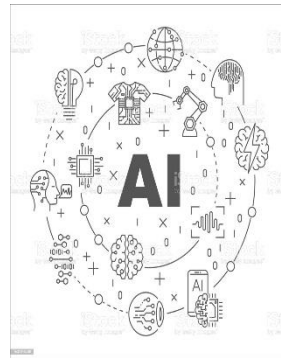
틀린 문항의 원인은 단계적으로 분석하여 부족
한 부분을 채우는 맞춤 처방

취약 개념 마스터

빅데이터 분석으로 발견된 취약 개념을 재학
습하여 완벽 마스터

기억 관리

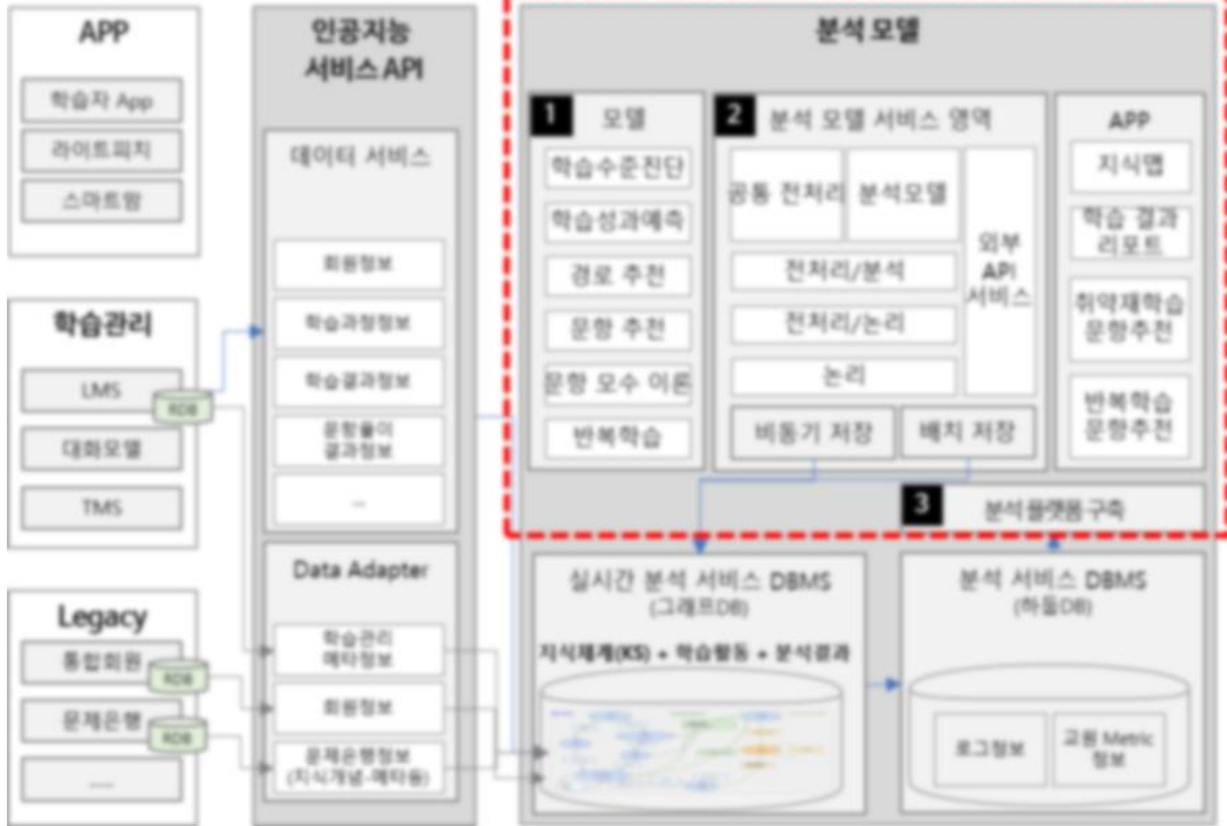
개인별 망각 주기를 분석하여 잊혀지는 개념
을 장기기억으로 완성



DBDISCOVER



서비스 영역



서비스구성요소

- 1 분석 모델(6개)**
 - 학습자 수준 진단(Bayesian Knowledge Tracing)
 - 학습 성과 예측(다중회귀분석)
 - 경로 추천(연관분석)
 - 문항 추천(협업 필터링)
 - 문항 추천(IRT)
 - 반복 학습(HLR)
- 2 분석 모델 서비스 영역**

| 프로그램 | 수량 |
|------------|-----|
| 공통 전 처리 영역 | 8개 |
| 분석 모델 영역 | 8개 |
| 기타 공통 영역 | 33개 |
| API 서비스 영역 | 27개 |
| 저장 영역 | 10개 |
| 총계 | 86개 |
- 3 분석 플랫폼 구축**
 - 분석 플랫폼 구축을 위한 하둡 활용 분석 스크립트 개발

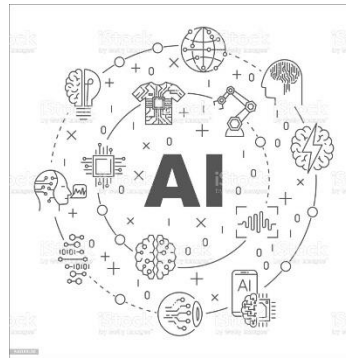
- 교육 인공지능 부분을 기획, 모델 및 서비스를 개발하여 취약 부분 예측, 성적 예측 등의 서비스를 제공

분석
 학습 성취도 분석
 학습 취약점 분석

예측
 성취도 기반 석차 예측
 출제 경향 별 문항 예측

추천
 맞춤 커리큘럼 추천
 취약점별 문항 추천

트레이닝
 취약 개념 집중 트레이닝
 선택 주제별 집중 트레이닝 성



DBDISCOVER



End of Document

김정일

jungkim720@DBdiscover.com

DBDISCOVER

transforms Data Into Business

www.DBdiscover.com