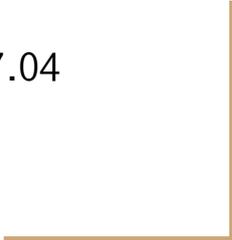




딜레마

딜러닝 서비스 상용화의 딜레마

2 nd DLCAT 2019.07.04
송규예



Deep learning + Dilemma
= Deep-lemma

: 딥러닝 기술기반 서비스의 진퇴양난. 궁지

여기 한 사업가가 있습니다
그의 1분 스피치를 들어봅니다

A person wearing a dark blue suit jacket and a yellow and blue striped tie is leaning over a light-colored wooden table. Their hands are resting on the table surface. In the center of the table, there is a yellow rectangular nameplate with the letters 'CEO' in bold black font. A semi-transparent dark grey rectangular box is overlaid on the image, containing Korean text.

안녕하세요. 이 솔루션은 딥러닝을 적용하여 비용을 절감하고,
더 나은 성능을 낼 수 있습니다.

CEO



요즘 비즈니스 하는데 인공지능 적용안하면 어디서 명함도 못내민다잖아.
정부과제도 다 '인공지능'을 꺼서 해야하고. 인공지능, 4차산업혁명. 필수지



그래 우리 회사도 인공지능을 도입해보자!
인공지능을 적용하면 투자 받기도 좋고 사업에도 도움이 되겠지

투자자들은 인공지능을 하는 회사에 투자하는 걸까요?

투자자들은 인공지능을 하는 회사에 투자하는 걸까요?



투자자들은 인공지능을 하는 회사에 투자하는 걸까요?

돈 잘 버는 회사

그럼 돈을 벌지 못하는 기술 회사들은 왜 투자를 받나요?

그럼 돈을 벌지 못하는 기술 회사들은 왜 투자를 받나요?

누군가는 그 기술을 비싸게 살 것이기 때문에

대표

우리회사도 인공지능을 적용할겁니다.
딥러닝으로 고객도 모으고 서비스 성능도 올리면 좋을테죠



머신러닝 엔지니어

흠..서비스에 딥러닝을 적용하실
거면 **딥레마**를 고려해야 합니다

머신러닝 모델을 돌려보고 실험하고
논문으로 정리하는 연구 프로세스



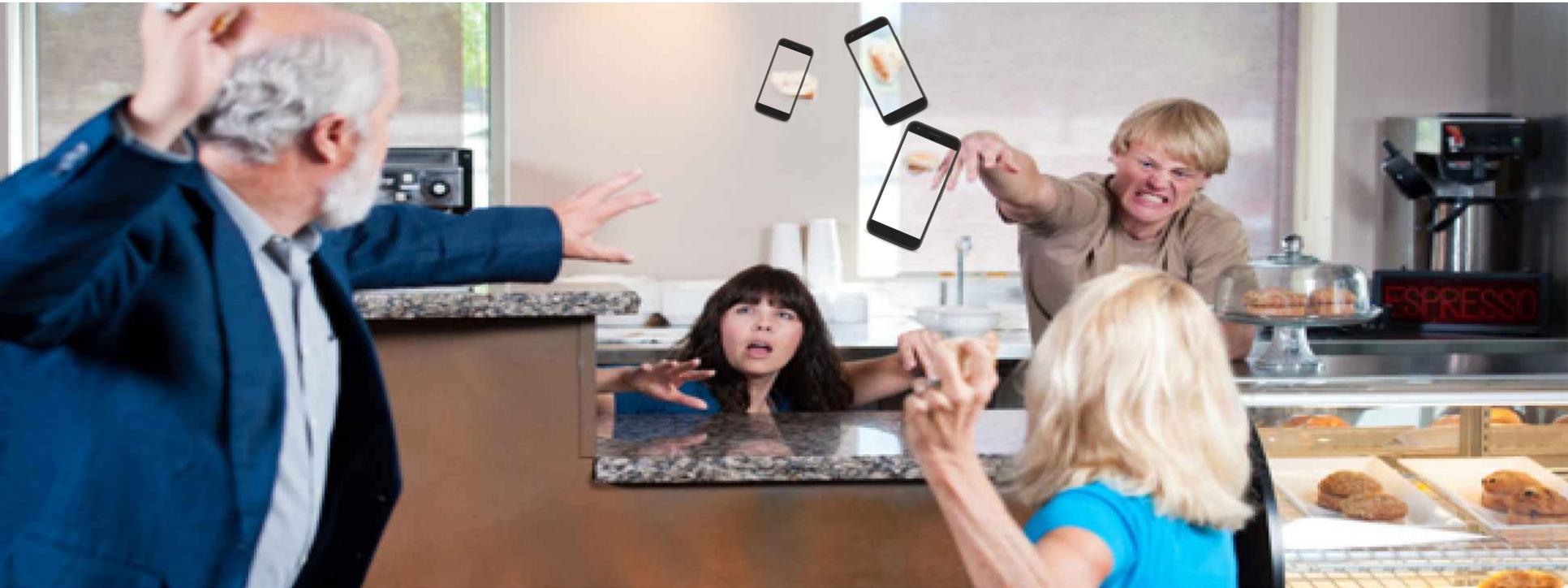


오호. 이런 인공지능 서비스가 있었네?
어디 한번 사용해볼까



흠.. 왜이렇게 느리고
사용하기가 불편하지? 이게 인공지능?

아오. 이게 서비스냐??
비싸기만하고 별로야!





고객 및 사용자 입장에서.
그리고 사업가 입장에서.
딤러닝기반 서비스



Caffé bene
gelato & waffle



머신러닝 측면

비즈니스 측면

서비스 상용화

머신러닝 기술 서비스 상용화는
기술 이상으로 고려해야할 부분들이
존재합니다

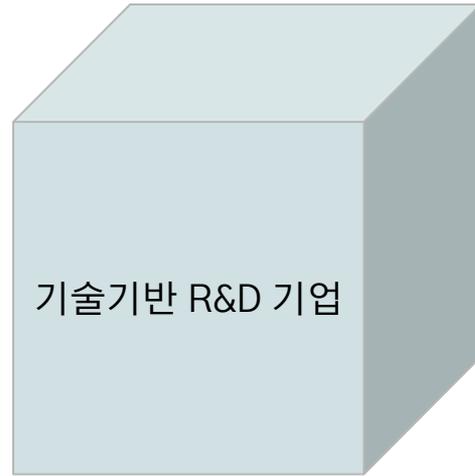
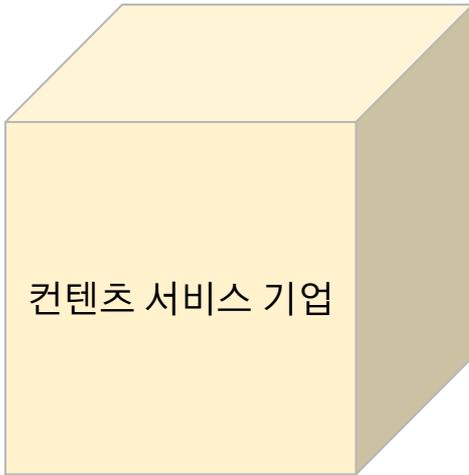
그렇다면 사업 및 서비스 상용화에 딥러닝을
적용하는 것에 어떤 고려사항이 있고
어떤 딜레마가 있을까요



발표자 소개



- (주) resonance 대표
- (주) 스피어웍 수석연구원
- HCI, 컴퓨터공학 전공
 - 머신러닝 엔지니어
 - 사업기획, 운영
 - 사용자 경험분석



컨텐츠 서비스 기업 측면

풀고자 하는 문제, 목표설정

시장성, 수익성이 있는가

사용자 및 고객 유입이 효과적인가

딥러닝 적용시 기술적 장점이 있는가

컨텐츠 서비스 기업 측면

풀고자 하는 문제, 목표설정

시장성, 수익성이 있는가

사용자 및 고객 유입이 효과적인가

딥러닝 적용시 기술적 장점이 있는가

이게 최선인가?
반드시 딥러닝이어야하나?

기술기반 R&D 기업 측면

풀고자 하는 문제, 목표설정

비용절감이 가능한 기술적 가치가 있는가

실제로 기술을 구현 가능한가

딥러닝 적용시 기술적 장점이 있는가

기술기반 R&D 기업 측면

풀고자 하는 문제, 목표설정

비용절감이 가능한 기술적 가치가 있는가

실제로 기술을 구현 가능한가

딥러닝 적용시 기술적 장점이 있는가

이게 최선인가?
반드시 딥러닝이어야하나?

기술기반 R&D 기업 측면

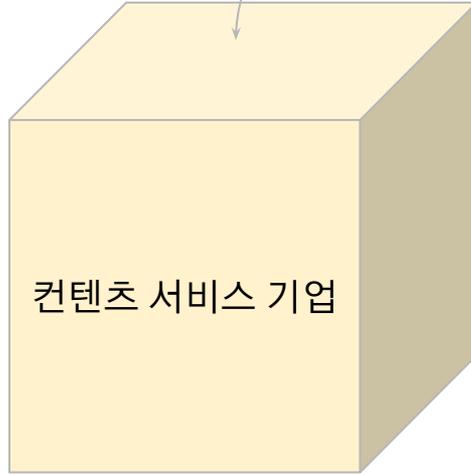
비용절감이 가능한 기술적 가치가 있는가

만약 전세계 유일무이한 기술을 가진 R&D 기업이 있다면 ‘비용절감’을 기술적 가치로 볼 수 있을까?

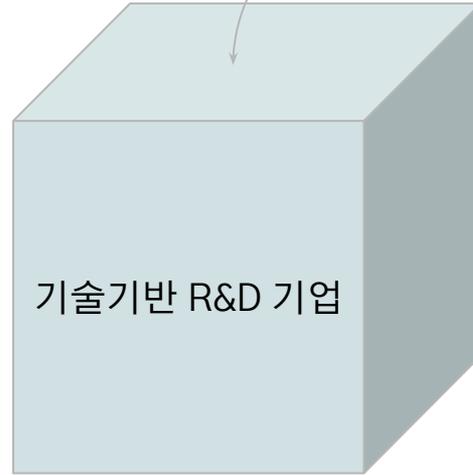
- 존재하지않는 기술 = 희소성이 무한대 = 가격이 무한대
- 존재하지않는 기술 ⇒ 존재하는 기술로 구현함으로써
가격을 무한대 ⇒ 유한대로 절감한 효과를 냄

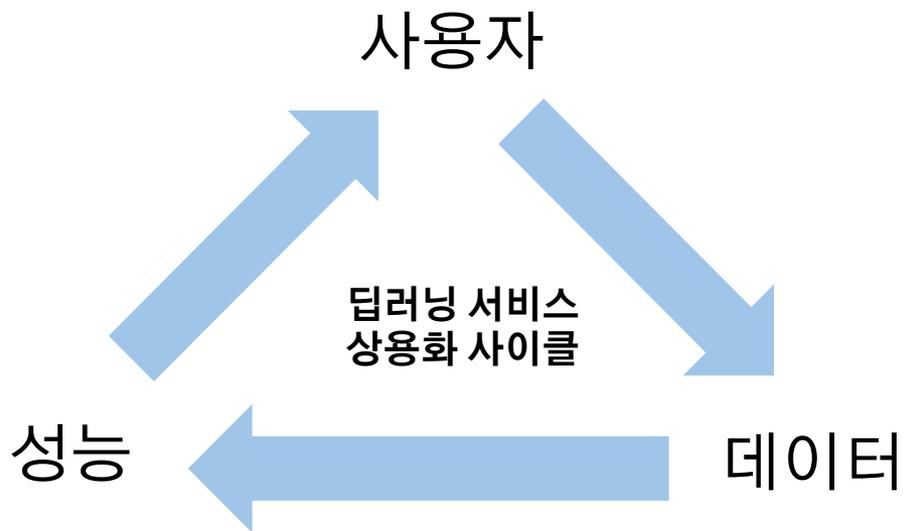
즉, 전세계 유일무이한 기술은 비용절감에 있어 기술적 가치가 있다고 볼 수 있다

딥러닝 기술



딥러닝 기술





사용자 부족

딤레마 사이클

성능 부족

데이터 부족

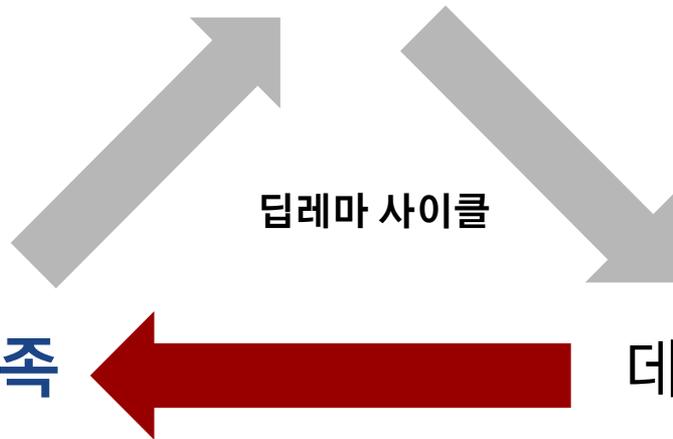


사용자 부족

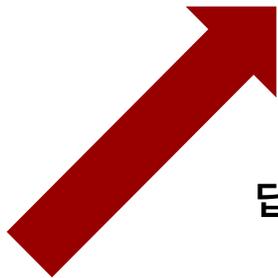
딤레마 사이클

성능 부족

데이터 부족



사용자 부족

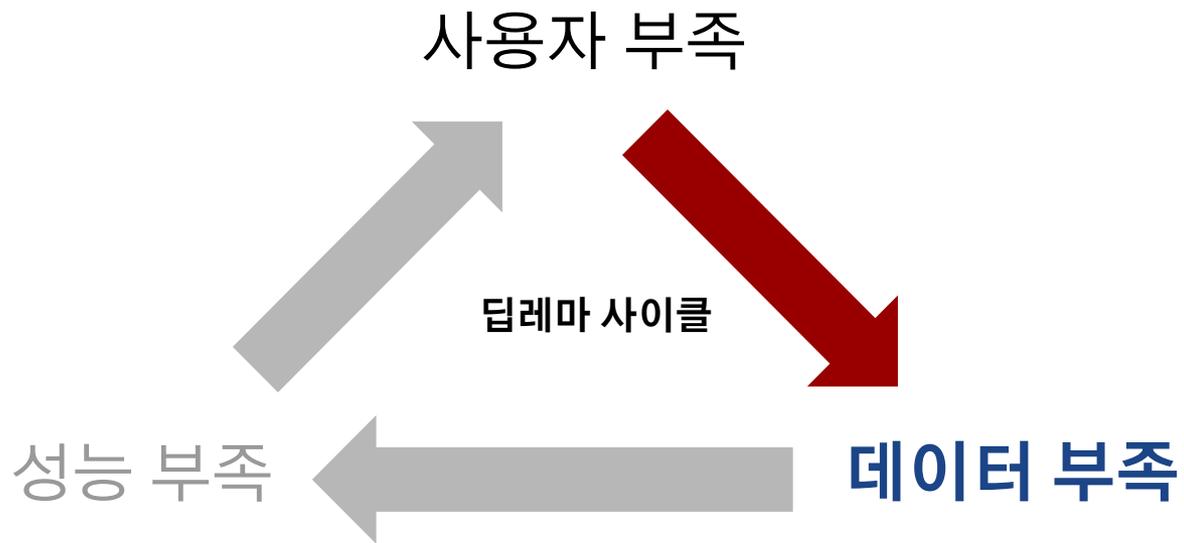


딤레마 사이클

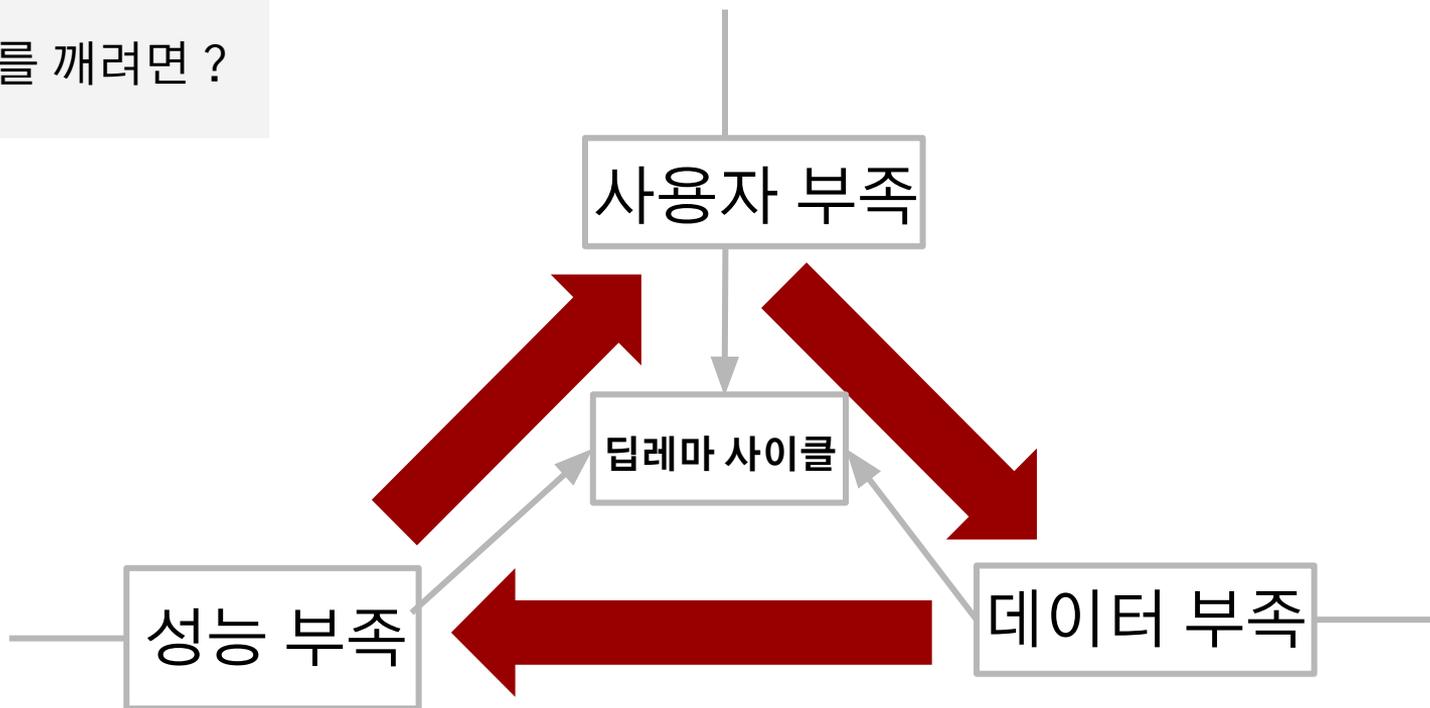
성능 부족



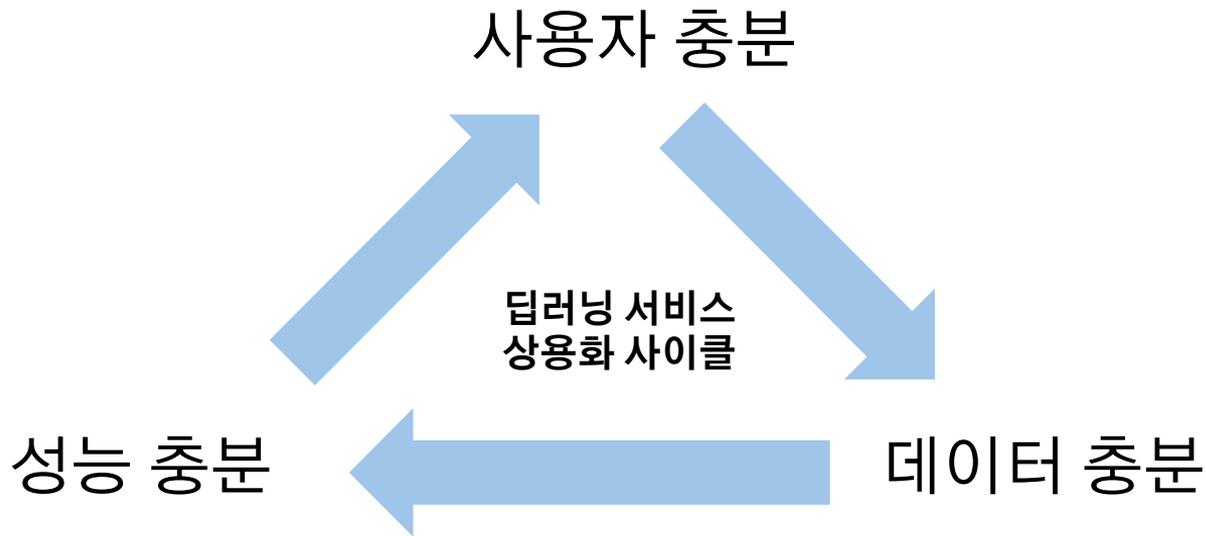
데이터 부족



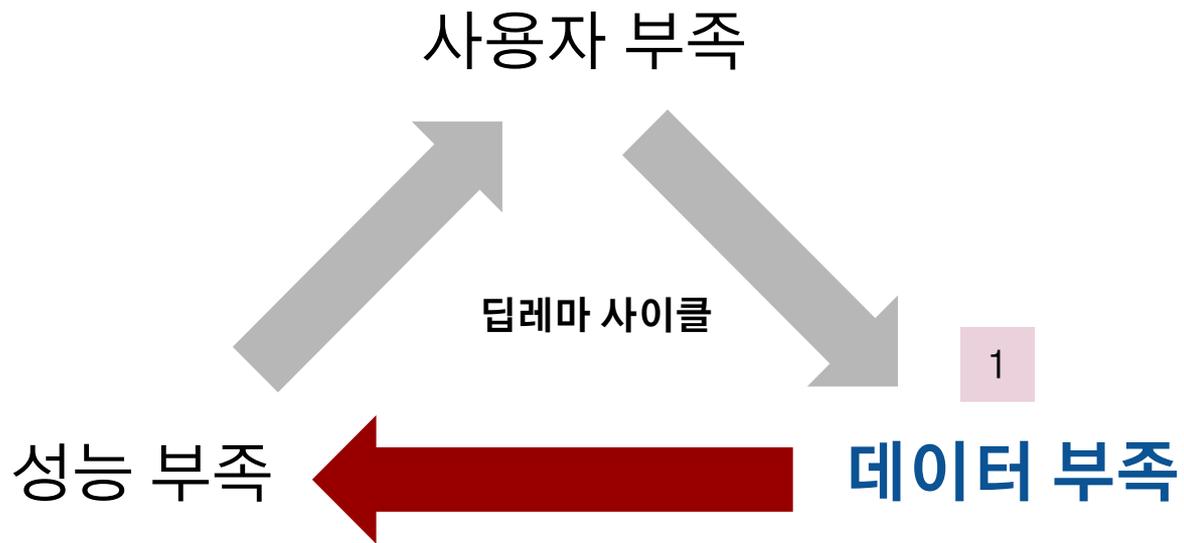
딤레마를 깨려면 ?



셋 중 어느 한 부분이라도 공략하여 부족에 의한 딤레마 고리를 깨야



올바른 딥러닝 서비스 상용화 사이클이 돌아감



서비스를 위한 딥러닝 모델을 개발할 때



..어음..일단 딥러닝 모델 개발을 위한 데이터 측면에서 보겠음...

서비스를 위한 딥러닝 모델을 개발할 때

일단 서비스에 적합한
머신러닝 **모델**을 넣고



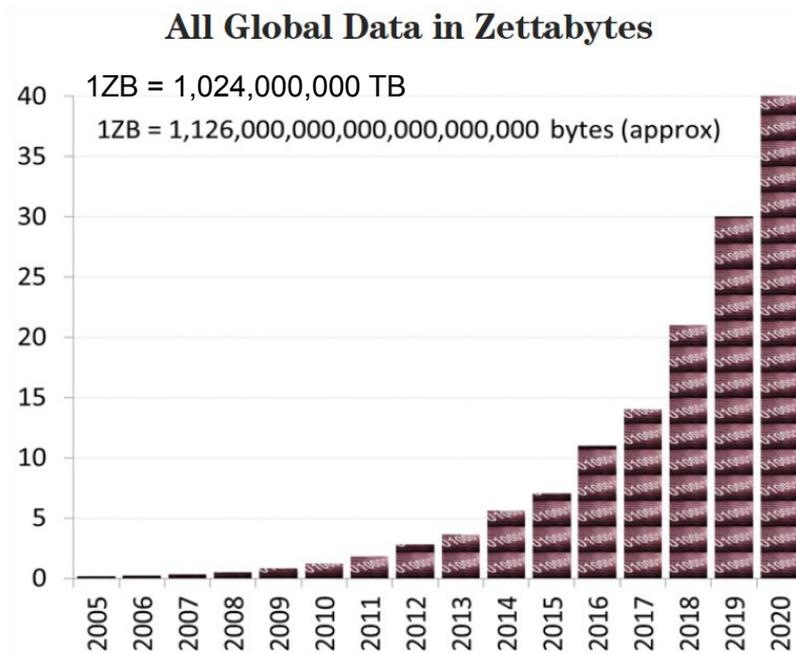
얼마나 길어질지 모르는
긴 실험 과정을 한 스푼 넣고



학습을 위한 **질 좋은 데이터**를..
넣었...아.. 꽤나 많이 필요하네;;;



데이터로 성능 올리기



Source: <http://www1.unece.org/stat/platform/display/msis/Big+Data>

???: 아니 데이터가 이렇게 많은데 왜 자꾸 데이터가 없다고 그러는거야?

???: 전 세계에 있는 데이터를 다 학습 시키면 되겠네!

???: 저거 뭐 다 크롤링 하면 되는거잖아?

???: 데이터에 돈이 왜 들어가???

1

데이터로 성능 올리기

딥러닝에 필요한 데이터

\neq Raw data

\equiv preprocessed train data

데이터로 성능 올리기

데이터를 모아 성능을 올리는 방법

(1) 오픈데이터(benchmark)를 활용하여 성능을 올린다

(2) 데이터 크롤링(collector)으로 성능을 올린다

(3) 제품(service)을 통해 데이터를 모아 성능을 올린다

1

데이터로 성능 올리기

항목	benchmark	collector	user service (UX)
데이터 개수	제한적	매우 큼	많아질 수 있음
정제 비용	없음	매우 큼	UX에 의존
개발비용	없음	큼	큼
퀄리티	높음	더러움	덜 더러움

데이터로 성능 올리기

(1) 오픈데이터(benchmark)를 활용하여 성능을 올린다

오픈 데이터(benchmark) = 양질의 상태이나 타겟이 정확하지않은 일반적이고 적은 양의 데이터

현실 데이터(wild data) = 가공이 필요한 상태 (편향존재, 후처리 필요)



딥러닝 기반 사용자 맞춤형 서비스를 하려면 많은 양의 목표에 적절한 오픈데이터가 필요하며

현실 데이터로 보충하기엔 추가적인 가공이 필요



오픈데이터 활용의 한계로 인해 사용자 맞춤은 커녕 general하게 적용되는 서비스를 할 수밖에 없음

이전에 쓰던 rule-based가 더 좋은 성능을 보일수도

1

데이터로 성능 올리기

(2) 데이터 크롤링(collector)으로 성능을 올린다

성능을 올리기 위한 크롤링 진행 시 많은 양의 질 좋은 데이터를 수집 및 가공해야함



데이터 수집, 가공 인프라를 구축하고 진행하는 것에 대부분의 자본이 들어갈 수밖에 없음



정작 딥러닝 모델을 통한 서비스를 진행할 시, 여유자금이 없음

배보다 배꼽이 더 클수도

데이터로 성능 올리기

(3) 제품(service)을 통해 데이터를 모아 성능을 올린다

자체 제품을 통해 사용자를 모아 데이터를 수집할 수 있음



페이스북, 구글, 삼성처럼 회사를 키워 많은 양의 데이터를 수집 (...)

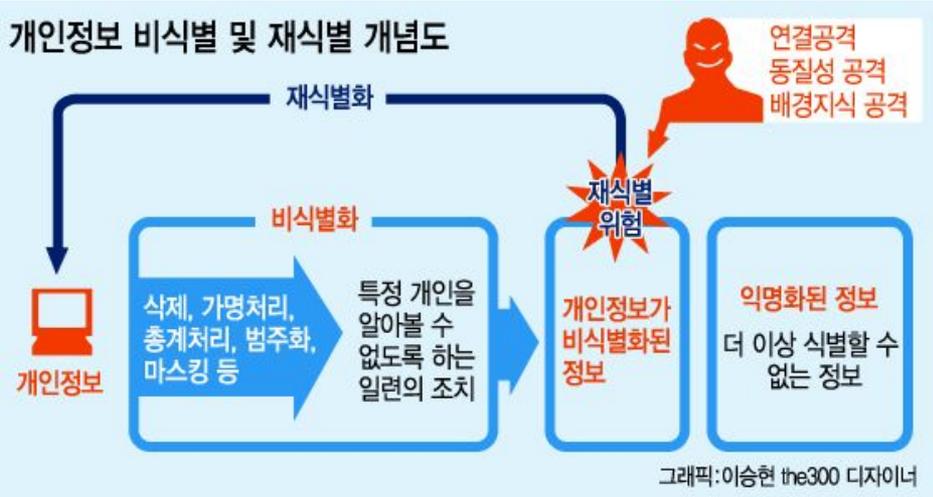


보통 대기업처럼 만들기는 당장은 힘드므로 정부사업, 데이터 바우처 협력을 통해 데이터를 수급받음

데이터로 성능 올리기

(3) 제품(service)을 통해 데이터를 모아 성능을 올린다

딥러닝 서비스를 위한 데이터 수집 시 : 데이터 보안 문제 고려하기



- 정보를 판다면서요? 9시간 전

☆☆☆☆☆

내용이 지워지지않고 개발자가 저장했다가 고민, 위치정보 등을 판매한다고 합니다 양심없으시네요
- 당신들의 고민이 곧 돈입니다 9시간 전

☆☆☆☆☆

익명보장을 담보로 모은 연령대, 지역, 성별, 주로 하는 고민등을 차곡차곡 모아 데이터베이스업체에 팔아버리기~
- 데이터를 팔아요? ㅎㅎ 12시간 전

☆☆☆☆☆

익명성을 없애는 내용을 삭제하고 그 데이터를 판다... 아주 법적으로 문제 하나 없고 좋네요 아주? 어이가 없어서; ㅎㅎ 빨리 데이터 올린거 삭제하고 이용자들에게 사과하세요. 지우개가 아니라 하드디스크였네 완전!!!!

데이터로 성능 올리기

딥러닝 서비스를 위한 데이터 수집,가공,분석 시 : 인프라 관련 비용문제 고려하기

일단 서비스에 적합한
머신러닝 **모델**을 넣고



얼마나 길어질지 모르는
긴 실험 과정을 한 스푼 넣고



학습을 위한 **질 좋은 데이터**를..
넣었...아.. 꽤나 많이 필요하네;;;



데이터로 성능 올리기

딥러닝 서비스를 위한 데이터 수집, 가공, 분석 시 : 인프라 관련 비용문제 고려하기

일단 서비스에 적합한
머신러닝 **모델**을 넣고



얼마나 길어질지 모르는
긴 실험 과정을 한 스푼 넣고



학습을 위한 **질 좋은 데이터**를..
넣었...아.. 꽤나 많이 필요하네;;;





RDBMS



DB 인프라 데이터 파이프라인 개발



데이터 관리소프트웨어라벨링 툴



라벨링 툴
어노테이션 관리
프로그램
(검수자)



분산컴퓨팅



분산 시스템



병렬처리(GPU)



모델 개발



모델 배포 및 API 개발



DevOps 구축



무중단서비스배포
(CICD)



대용량 데이터
전송 인프라



annotation
인력관리



UI/UX기획



UI/UX 디자인



사용성평가



연산 최적화



....





RDBMS



DB 인프라 데이터 파이프라인 개발



데이터 관리소프트웨어 라벨링 툴



라벨링 툴
어노테이션 관리
프로그램
(검수자)



분산컴퓨팅



분산 시스템



병렬처리(GPU)



모델 개발



모델 배포 및 API 개발



DevOps 구축



무중단서비스배포
(CICD)



대용량 데이터
인프라



annotation
인력관리



UI/UX기획



UI/UX 디자인



사용성평가



연산 최적화



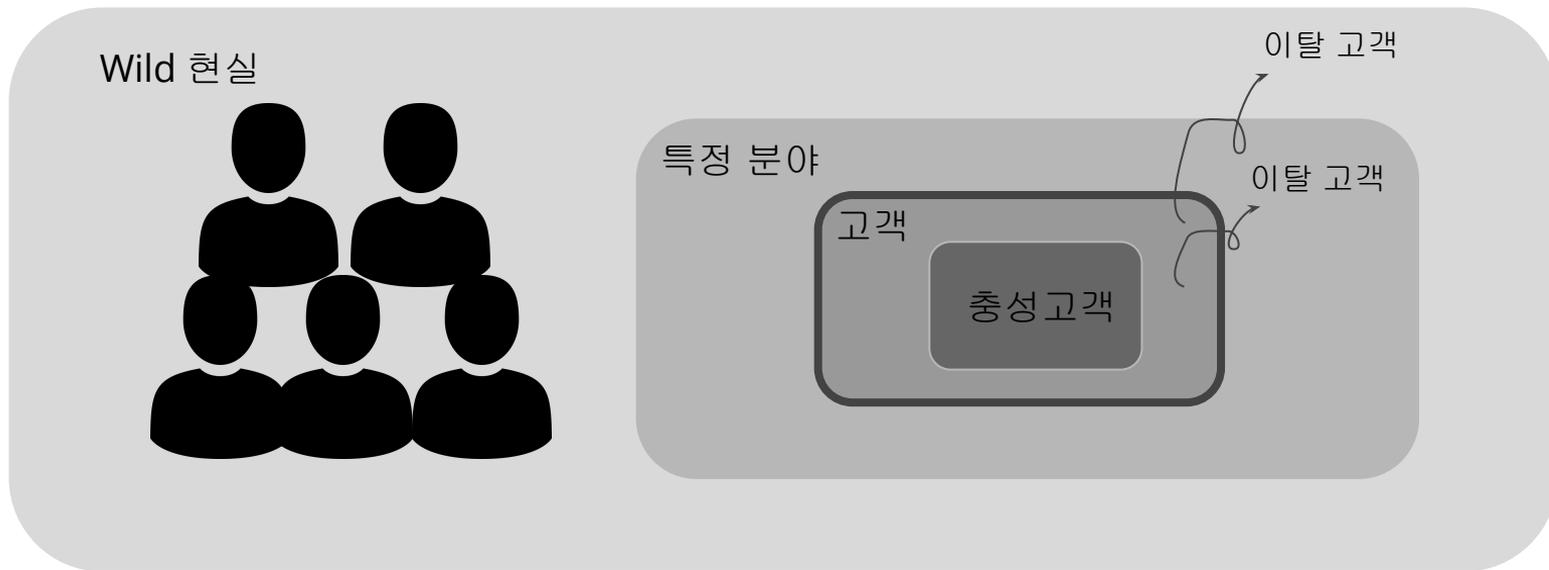
....



1

데이터로 성능 올리기

딥러닝 서비스를 위한 데이터 수집,가공,분석 시 : 데이터 편향, 미래예측 손실 고려





서비스를 위한 딥러닝 모델을 개발할 때



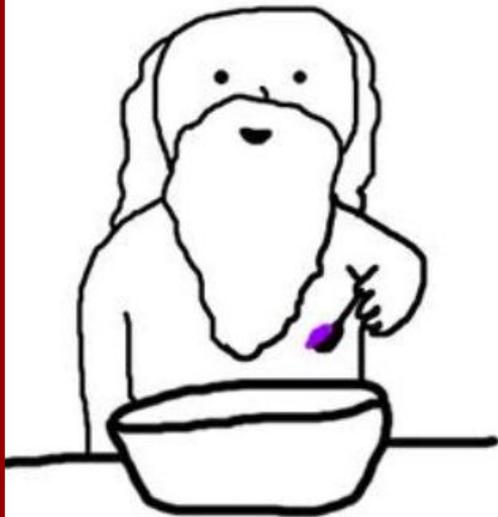
..어음..이번엔 딥러닝 모델 성능 측면에서 보겠음

서비스를 위한 딥러닝 모델을 개발할 때

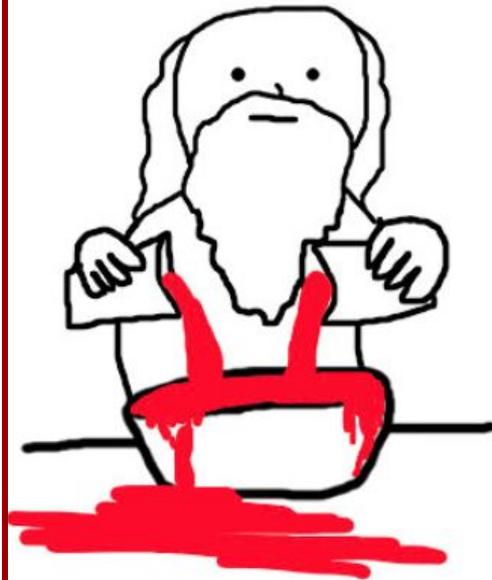
일단 서비스에 적합한
머신러닝 **모델**을 넣고



얼마나 길어질지 모르는
긴 실험 과정을 한 스푼 넣고



학습을 위한 **질 좋은 데이터**를..
넣었...아.. 꽤나 많이 필요하네;;;



성능을 올려 사용자 유입 증대

도메인 지식이 기반되어야 서비스에 적합한 기술성능을 구현할 수있다

Inspace, IBM왓슨 기업이 좋은사례 :
우주항공 및 각 도메인 지식보유,
problem-solving의 올바른 접근과정

풀고자하는 문제와 관련된 도메인 지식 필요



일반적인 경우 딥러닝을 사용하지 않고도
문제를 풀 가능성이 높음
(어설픈게 딥러닝 적용하면 어둠의 길로..)

풀고자하는 문제를 딥러닝으로 풀것인가 고민 필요



기술적 노하우를 쌓아 성능올리기
(최적화, 유지보수 포함)

성능을 올려 사용자 유입 증대

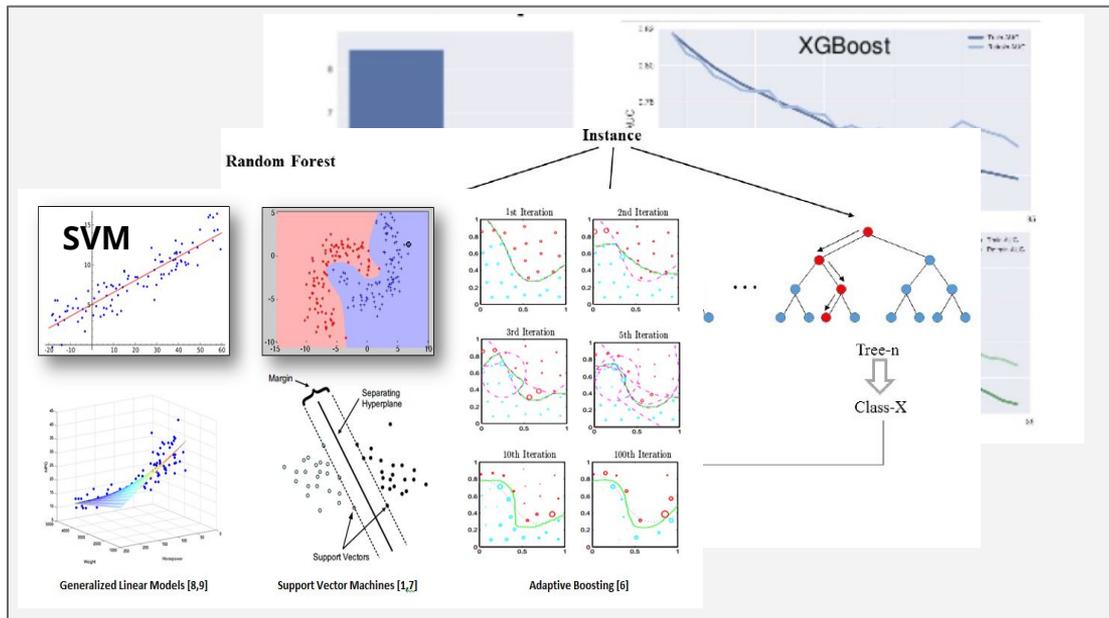
서비스 상용화를 생각한다면 딥러닝 모델만이 답이라고 할 수 없다

딥러닝 모델

많은 양의 데이터가 있어야 성능 확보
GPU 인프라, 분산처리 시스템 필요
블랙박스 요소 존재

고전적인 머신러닝 모델

데이터의 양이 상대적으로 적어도 가능
기존 시스템 인프라 활용가능
동작과정 명확, 튜닝에 용이



성능을 올려 사용자 유입 증대

서비스 기획시, 성능을 올리기 위한 긴 실험과정을 고려해야한다

데이터 수집 기간

시스템 인프라 작업

데이터 가공 기간

분석 모델 개발

데이터 학습

결과도출

모델 튜닝

최적화

성능을 올려 사용자 유입 증대

서비스 상용화에서는 기술적 정확도만이 답이 아닐 수 있다



데이터 분석과정을 자연스럽게 인지할 수있는 서비스 (심리테스트, 번역, 각종 분석결과앱)가 아닌
딥러닝 결과가 숨겨져 표현되는 서비스인 경우 즉각적인 반응속도가 중요.

성능과 다른 서비스 중요 요소들간의 trade-off 고려해야함

3

사용자 부족

딤레마 사이클

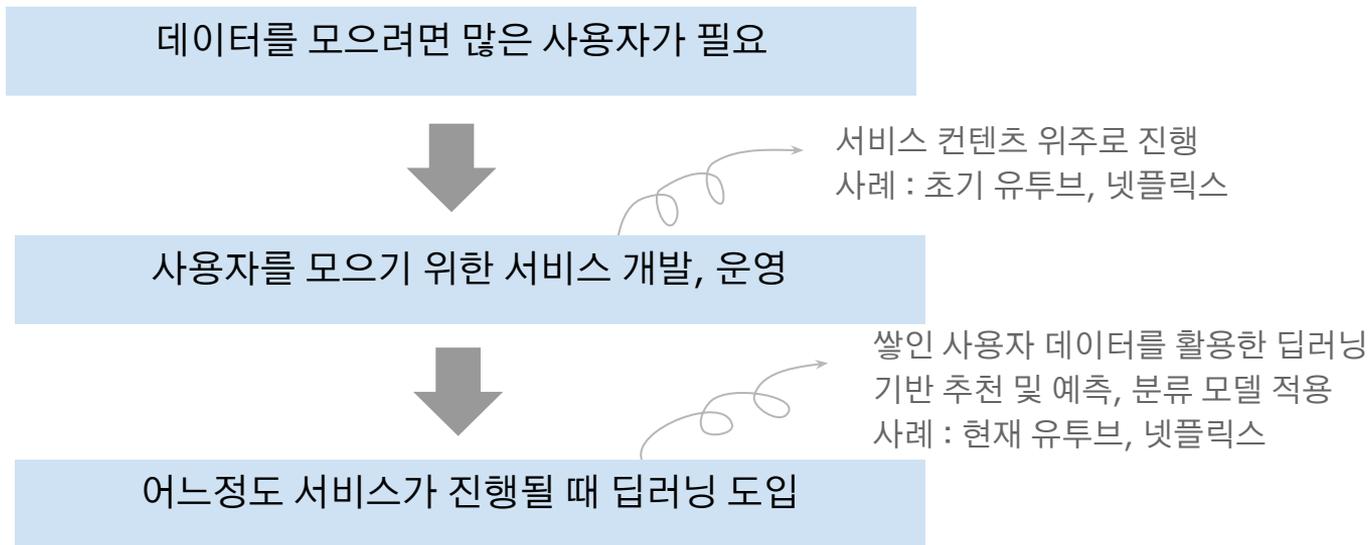
성능 부족

데이터 부족



사용자 유입증가로 데이터 증대

딥러닝 서비스를 위한 사용자 유입전략 : 콘텐츠, 서비스 개발



3 사용자 유입증가로 데이터 증대

딥러닝 서비스를 위한 사용자 유입전략 : UX 고려



사용자 유입증가로 데이터 증대

딤러닝 서비스를 위한 사용자 유입전략 : 변화에 따른 예측의 불확실성 고려



스컬프터(SCULPTOR) 룩북 ...
blog.naver.com



룩북-2015 LOOKBOOK여자...
akiii.co.kr



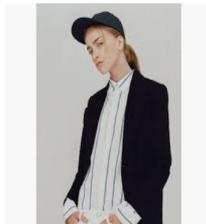
SALE FOR YOU : 광고 201...
saleforyou.egloos.com



SALE FOR YOU : 광고 201...
saleforyou.egloos.com



LOOKBOOK ; 브랜드룩북-LOOKBOOK ; 2...
fairplay142.com



사용자 유입증가로 데이터 증대

딥러닝 서비스를 위한 사용자 유입전략 : 사용자,사회문화에 따른 데이터 편향 고려



감정인식&표현



:(sad (슬픔)
%-) (혼란스러움)
:-D (행복함)

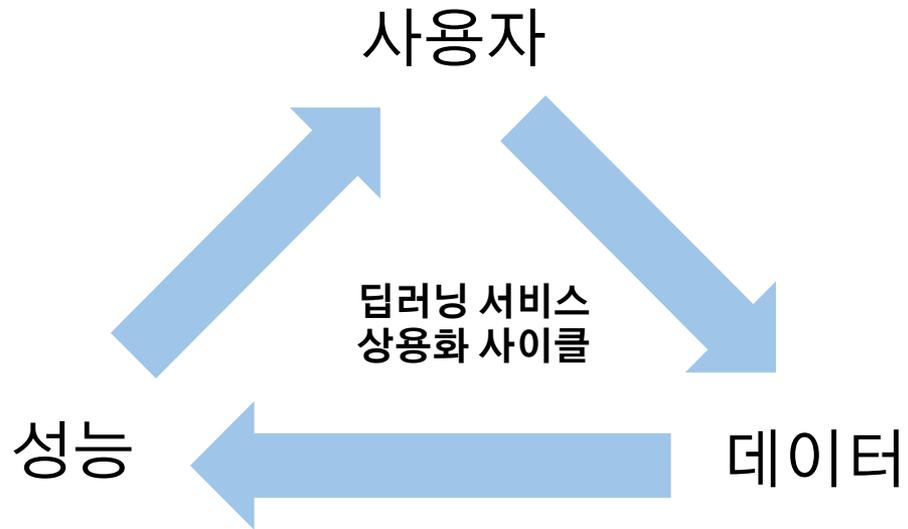


서양권



π_π (슬픔)
@_@ (혼란)
^0^ (행복함)

동양권

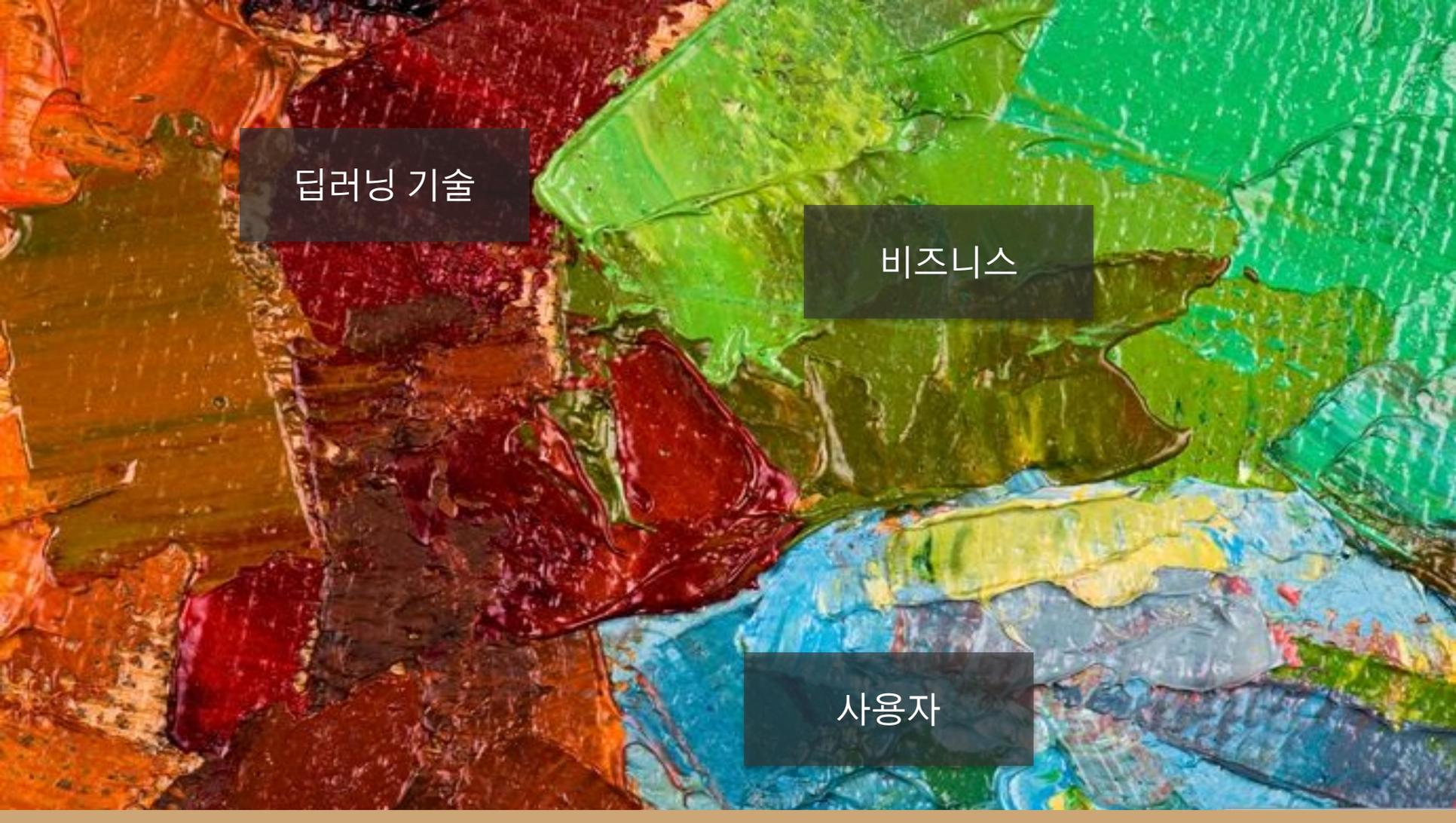




머신러닝 측면

비즈니스 측면

서비스 상용화

An abstract background featuring a complex, layered texture of various colors including deep reds, oranges, yellows, greens, and blues. The colors are blended and layered, creating a sense of depth and movement, similar to a marbled paper or a textured wall. Three dark grey rectangular boxes are overlaid on the image, containing Korean text.

딥러닝 기술

비즈니스

사용자



경청해 주셔서 감사합니다