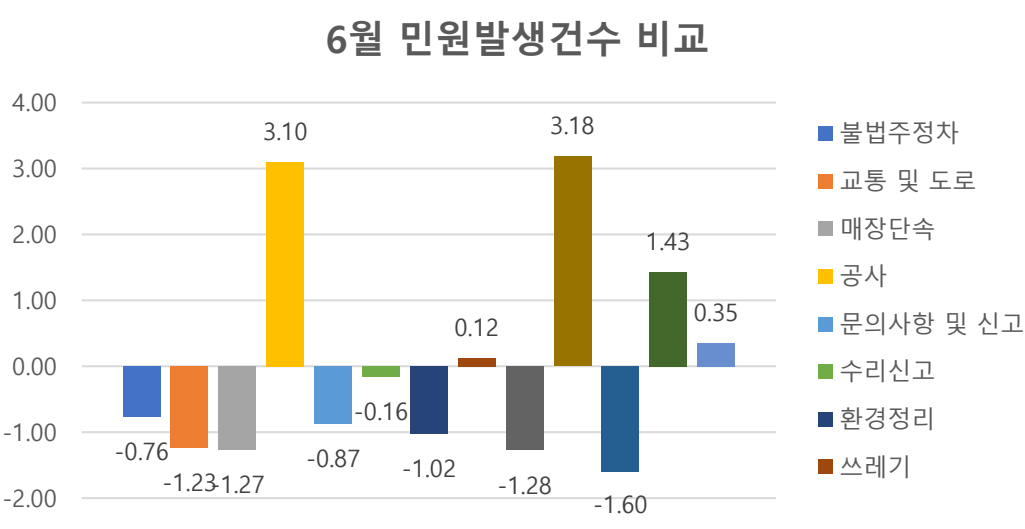
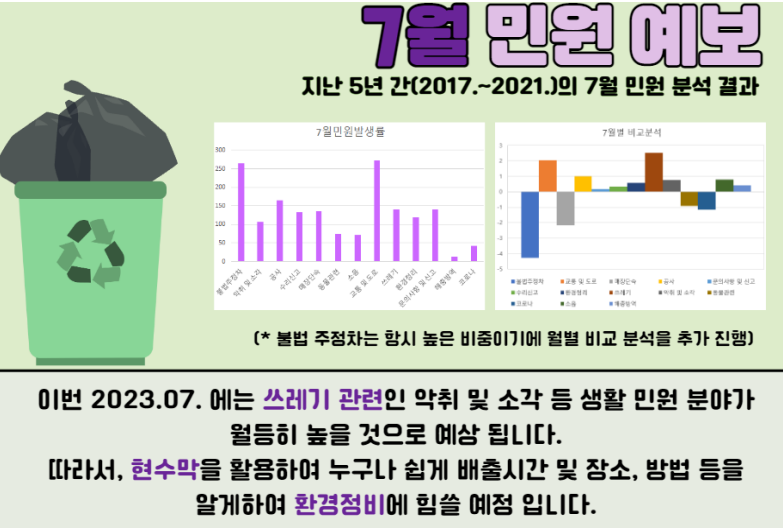


# Project 01

## 머신러닝을 활용한 민원 자동 분류 및 예측

- 연구 목적 : 유성구 민원 데이터를 통해 민원 자동 분류와 예측을 진행하고자 함
- 연구 방법 : 데이터 수집, 데이터 전처리, 민원 분류 모델 생성, 민원 통계 분석, 개선 방안 정책 제안
- 활용 기술 : Pandas, XGB, RF model, TF-IDF, DTM, Konlpy

민원번호	민원대분류	민원소분류	위치	위치키워드	민원내용
1	생활/시설민원	동물관련	원신흥동	원신흥동	원신흥동 어울림하트 뒷쪽 야산에서 지속적으로 야생동물 울음소리가 들려 주민들의 항의전화가 끊이지 않음.
2	환경민원	냄새	관평동	관평동	관평동 롯데마트 인근이며 가스냄새가 너무 심하게 남. 평소에도 냄새가 심하며 금일은 유독 냄새가 심함
3	주차민원	주차, 교통	노은동	반석역	반석역 공영주차장 근처 3층주차타워 근처 상습적 불법주정차로 교통혼잡야기, 평일야간도 단속 요청.

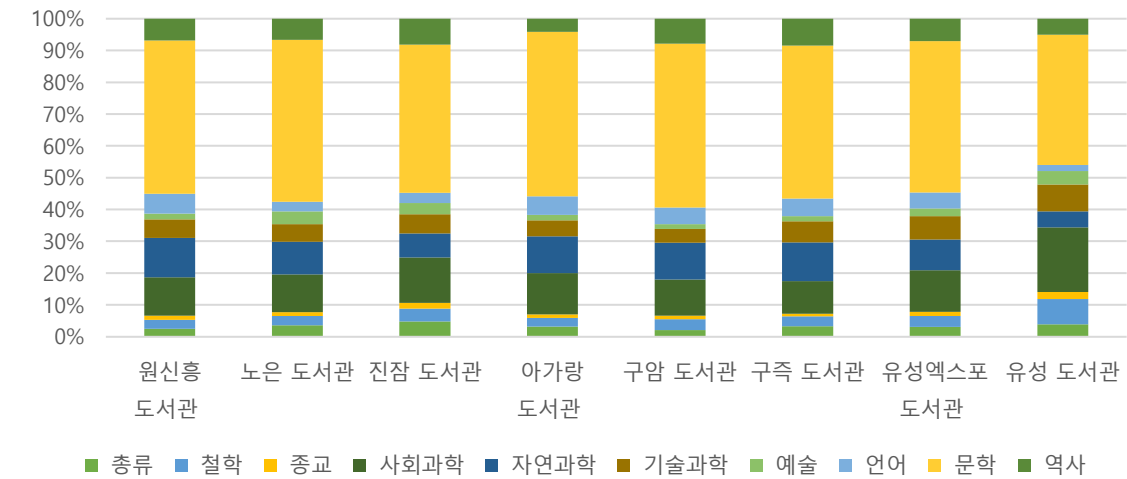


# Project 02

## 도서 수요 분석 데이터를 통한 도서 재분배 및 지식 콘텐츠 기획

- 연구 목적 : 유성구 도서 수요 데이터를 통해 도서관별 지식 수요 콘텐츠 기획 및 도서 재분배
- 연구 방법 : 데이터 수집, 데이터 전처리, 분류 모델 생성, 민원 통계 분석, 개선 방안 정책 제안
- 활용 기술 : Pandas, BeautifulSoup, K-NN, LSA, TextRank

장서/대출데이터				연기대출도서	도서별 이용분석	대출 급상승 도서	지역별 비교분석	어달의 키워드	참여 도서관
지역	대전	세부지역	유성구	도서관유형	전체	도서관명	도서관 선택	Search	구독신청
도서관명		데이터 유형		공개범위		최근 데이터 제공일			
관평도서관 장서/대출 데이터		Text Excel API		장서/대출		2023-05-01			
구암도서관 장서/대출 데이터		Text Excel API		장서/대출		2023-05-01			
구죽도서관 장서/대출 데이터		Text Excel API		장서/대출		2023-05-01			
노은도서관 장서/대출 데이터		Text Excel API		장서/대출		2023-05-01			
원신흥도서관 장서/대출 데이터		Text Excel API		장서/대출		2023-05-01			
유성도서관 장서/대출 데이터		Text Excel API		장서/대출		2023-05-01			
유성도서관 엑스포분관 장서/대출 데이터		Text Excel API		장서/대출		2023-05-01			
진잠도서관 장서/대출 데이터		Text Excel API		장서/대출		2023-05-01			



도서관	연령	성별	관심 분야
구죽도서관	10대	남	810 한국 문학, 740 영어
구죽도서관	20대	남	180 심리학, 580 제조업
구죽도서관	30대	남	130 철학의 세계, 070 신문

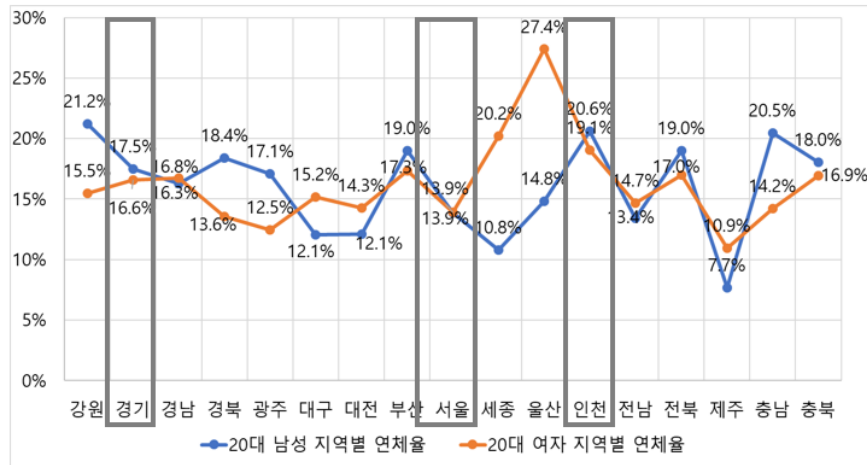
관심사	연령별	성별	문화 행사
000 총류	10대	남	“공부머리를 좌우하는 초격차 읽기능력의 비밀 초등메타인지 독서법”
100 철학	10대	남	“독서동아리 하브루타 생각논술 모집”
100 철학	20대	남	“니체에게 배우는 차라투스트라의 지혜”
100 철학	30대	남	“성인독서동아리(필사동아리) 회원 추가 모집”

# Project 03

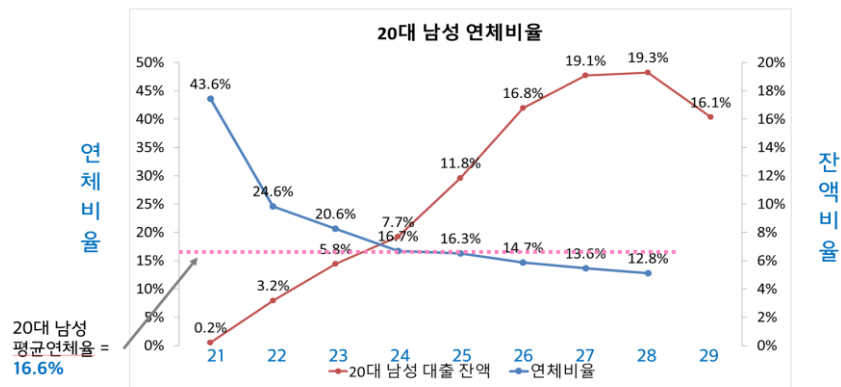
## 대출데이터를 통한 신규 유입자 확보 및 기존 고객 관리

- 연구 목적 : 한성저축은행의 소액대출 데이터를 통해 현재 고객에 대한 관리와 신규 고객 관리 방법 제시
- 연구 방법 : 데이터 전처리, 통계 분석, 개선 방안 정책 제안

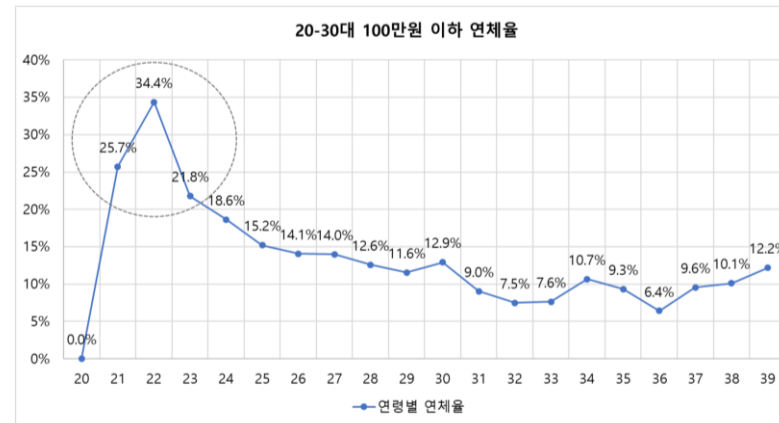
### 20대 고객의 지역별 연체율 및 대출잔액(2월20일기준)



### 20대 남성 고객의 연체율 및 대출잔액(2월20일기준)



### 연령별 100만원이하 대출 고객의 연체율(2월20일기준)

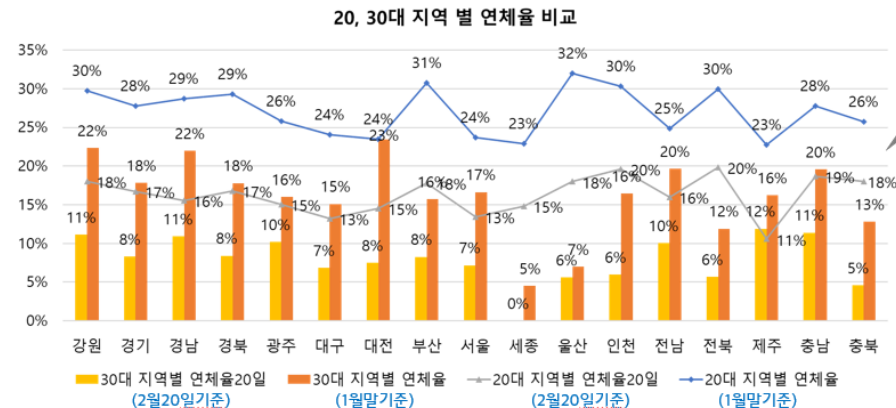


(2월20일 기준)

- 20대 초반(21, 22, 23)의 연체율 여전히 높음
- 20대 초반의 계좌수는 매우 낮아, 20대 초반의 특성 파악 필요

14.27%	전체 평균 연체율
13.16%	23세 이상 평균 연체율

### 20, 30대 고객의 지역별 연체율 비교(1월말기준 vs 2월20일기준)



- 20대의 지역별 연체율이 (1월말→2월20일) 많이 낮아졌으나 여전히 높음

42

31

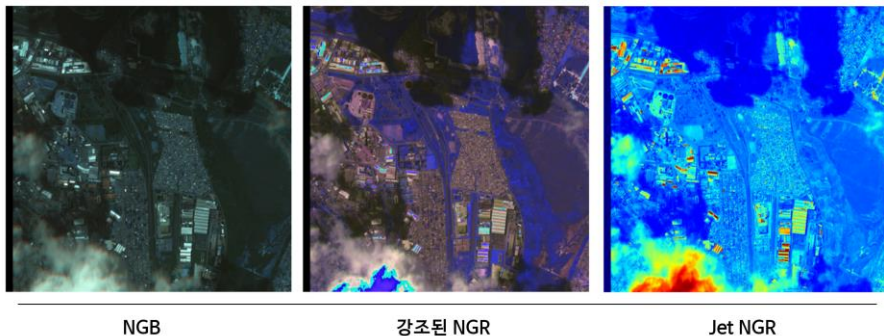
# Class 01

## 컴퓨터 비전 : Clouds Semantic Segmentation

- 주요 내용 : 인공위성 사진에서 구름을 추출 Semantic Segmentation 모델을 학습하여 인공위성 사진에서 구름 찾아 표시구름은 짙은 구름, 얇은 구름, 구름 그림자 총 3가지이며 각각 RGB 채널 값은 붉은색(255,0, 0), 녹색(0, 255, 0), 노란색(255, 255, 0)으로 표시
- 활용 기술 : Pytorch, torchvision models, efficient-net, Resnet, Unet 구조
- 참고 사이트 : <https://www.Kaggle.com/competitions/2023spring-clouds-segmentation/overview>

### 데이터 전처리(색상 강조)

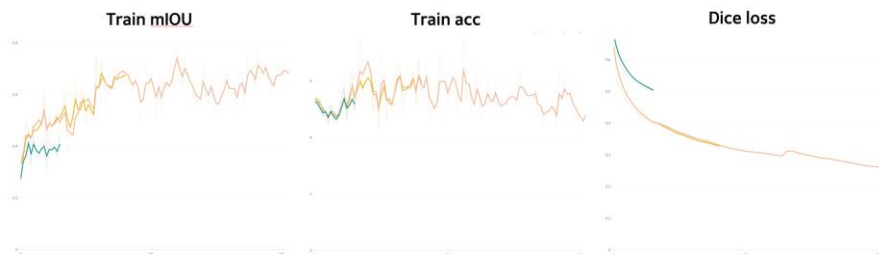
- 색상을 강조하면 학습을 더 잘할까?



### 데이터 전처리 비교

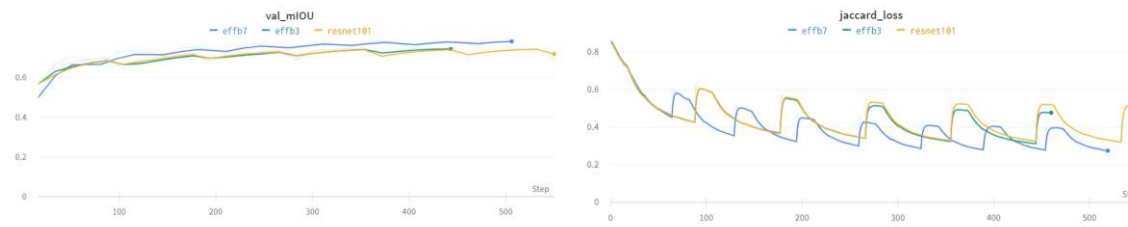
위에 세가지 전처리에 대한 -Train mIOU, Train\_acc, dice\_loss 값 비교

- 초록색 (RGB-NGR)
- 살구색 (RGB+NGR)
- 노란색 (RGB+NGR, 색상 강조)



### 모델 성능비교

- Deeplabv3\_resnet101 Public Score **0.82819**
- Deeplabv3\_resnet101 pretrained Public Score **0.80356**
- Unet구조에서 efficient-net b3으로 진행 Public Score **0.83573**



### Scheduler 적용

Baseline scheduler를 사용함 1에폭은 너무 오래 걸려서,

500 step마다 validation을 진행하고 learning rate 수정하도록 함

Patience는 5로 설정함 (6600 step정도 진행 후에 learning rate가 절반으로 감소함)



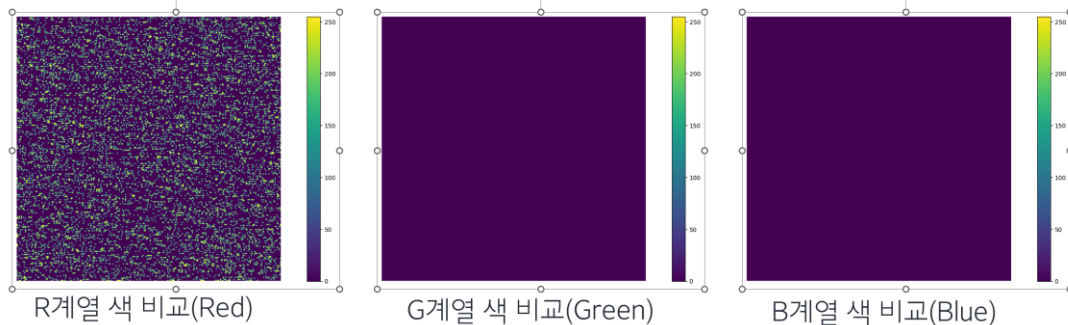
## Class 02

# 인공지능 : Image Anomaly Detection

- 주요 내용 : 사진을 원본 이미지(cover)인지 메시지가 삽입된 이미지(stego)인지 판별하는 과제 Image Classification 모델을 학습하여 각 사진을 cover image 인지 stego image 인지 분류하는 문제
- 활용 기술 : GroupKFold, albumentations, pytorch, efficient-net
- 참고 사이트 : <https://www.kaggle.com/competitions/image-anomaly-detection>

cover, stego 이미지 비교

```
difference = np.abs(cover_image - stego_image)
```



Pre-trained 모델에서 4개의 정답 존재

이미지 크기 512\*512

Convert RGB/BGR to Luma-chroma

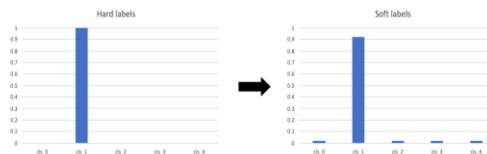
$$V \leftarrow \max(R, G, B)$$
$$S \leftarrow \begin{cases} \frac{V - \min(R, G, B)}{V} & \text{if } V \neq 0 \\ 0 & \text{otherwise} \end{cases}$$
$$H \leftarrow \begin{cases} 60(G - B)/(V - \min(R, G, B)) & \text{if } V = R \\ 120 + 60(B - R)/(V - \min(R, G, B)) & \text{if } V = G \\ 240 + 60(R - G)/(V - \min(R, G, B)) & \text{if } V = B \\ 0 & \text{if } R = G = B \end{cases}$$

레이블 스무딩

One-hot encoded vector를 Soft label로 스무딩

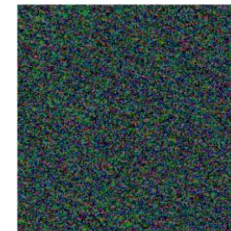
$$\mathbf{y}_k^{\text{LS}} = \mathbf{y}_k(1 - \alpha) + \alpha/K$$

K : Class의 수,  $\alpha$  : 0.05

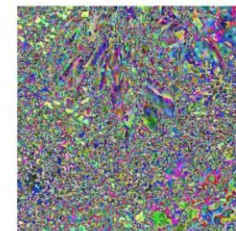


## 실험결과분석

정성적 분석 - 스테가노그래피 복호화



Hidden bits : 1



Hidden bits : 5



Hidden bits : 7



# Training 01

## 서울 ICT 이노베이션 스퀘어 : 고급 언어인지 과정 수료(160H)

- 주요 내용 : 자연어 처리에 대한 이해와 Word Embedding, KNN, CNN, Classification, LM에 대한 이해를 함
- 활용 기술 :

제 AI-22-22-0007366 호

### 수료증

성명 : 나종열

생년월일 : 1999.06.18.

과정구분 : 인공지능 고급

과정명 : AI 기술활용(고급언어과정)

교육기간 : 2022.02.12 ~ 2022.04.17 (160시간)

위 사람은 정보통신산업진흥원에서 시행한  
인공지능 교육 상기 과정을 이수하였기에  
이 증서를 수여합니다.

1주차	자연어처리 개요	5주차	Language Model
2주차	Word Embedding	6주차	Attention
3주차	RNN, CNN, Classification	7주차	SPAM분류, POS Tagging
4주차	Text Similarity	8주차	NER 모델

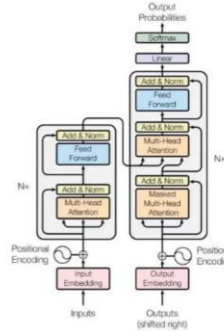
2022년 04월 17일

정보통신산업진흥원장



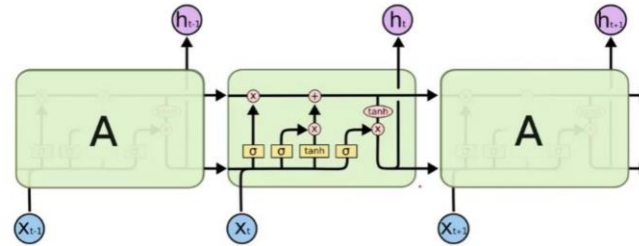
#### 트랜스포머 Transformer

- 주의 메커니즘만을 사용한 Seq2Seq 모형
- 문장 내의 주의 메커니즘 적용
- 문장 간에도 주의 메커니즘 적용



#### 순환신경망과 주의 메커니즘 : Long Shot Term Memory

- 순환신경망의 사라지는 경사 문제를 해결



#### 데이터 인코딩

##### 레이블 인코딩

[TV, 냉장고, 전자렌지, 컴퓨터, 선풍기, 믹서] → [0, 1, 4, 5, 3, 2] `encoder.fit(items)`

##### 원본 데이터

상품 분류	가격
TV	1,000,000
냉장고	1,500,000
전자렌지	200,000
컴퓨터	800,000
선풍기	100,000
선풍기	100,000
믹서	50,000
믹서	50,000

##### 상품 분류를 레이블 인코딩 한 데이터

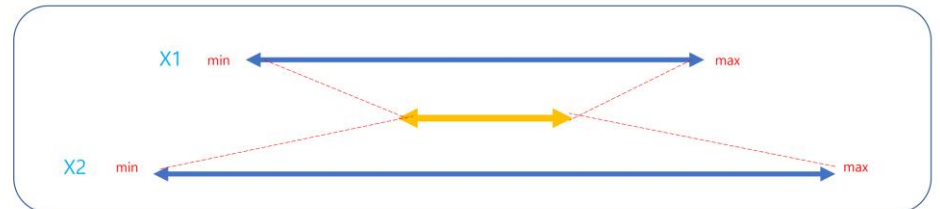
상품 분류	가격
0	1,000,000
1	1,500,000
4	200,000
5	800,000
3	100,000
3	100,000
2	50,000
2	50,000

`labels = encoder.transform(items)`

#### 피쳐 스케일링(feature scaling)

##### 피쳐 스케일링(feature scaling)

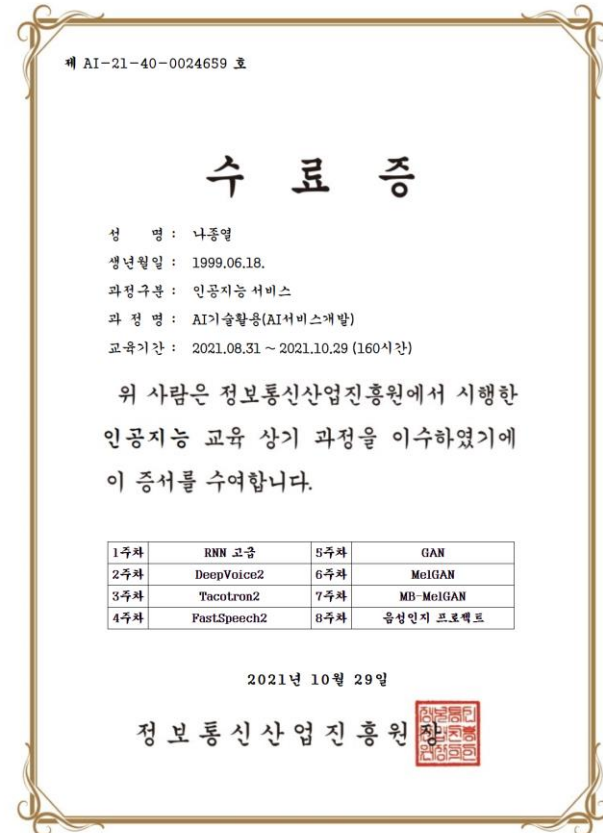
$$\hat{y} = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p + \epsilon$$



# Trainning 02

## 서울 ICT 이노베이션 스퀘어 : AI 서비스 기획 과정 수료(160H)

- 주요 내용 : 고객의 프로파일을 분석을 기반한 부동산 매물 매칭서비스 개발 및 기획



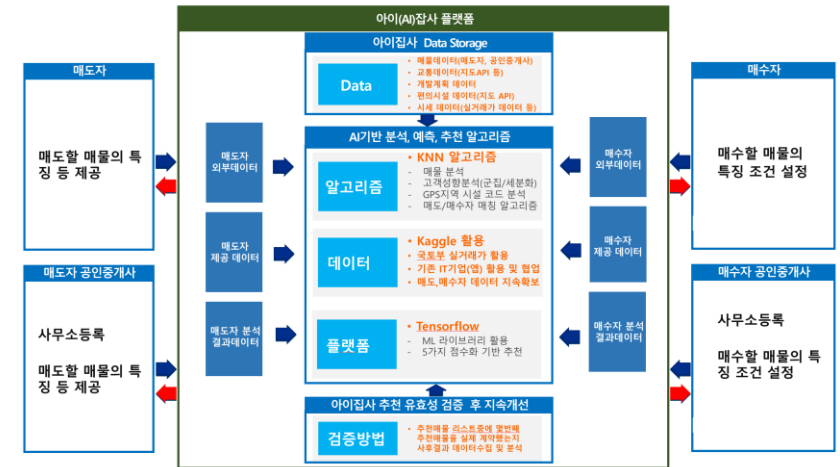
아이집사 BM Canvas

<p>핵심파트너</p> <p>공인중개사</p> <p>지도 API (카카오, 구글)</p> <p>감정평가사</p> <p>실거래가 정보(국토부)</p> <p>기존IT포털(네이버, 다음, 직방)</p> <p>데이터공공포털(별안부)</p>	<p>핵심활동</p> <p>고객 맞춤 AI 부동산 중개 서비스</p> <p>주변 편의시설 분석</p>	<p>가치제안</p> <p>1. AI가 추천하는 나의 집</p> <p>2. 검색시간과 중개 비용 절약</p> <p>3. 주변 시설에 대한 정보 확보</p> <p>4. 인공지능의 추천지수를 분석해 집수로 분석</p>	<p>고객관계</p> <p>매물 수 증가를 통한 회원 확대</p> <p>공인중개사 대상 광고지원</p> <p>분양 대행사</p> <p>공인중개사</p> <p>대출정보(KB은행)</p>	<p>고객 세그먼트</p> <p>중개용 사용자</p> <p>인테리어 업체</p> <p>관리업체(청소용품)</p> <p>분양 대행사</p> <p>공인중개사</p> <p>대출정보(KB은행)</p>
<p>데이터</p> <p>지도 API 주변 시설</p> <p>매물데이터</p> <p>개발, 계획</p> <p>교통</p> <p>편의시설</p> <p>사제</p>	<p>핵심자원</p> <p>부동산중개사</p> <p>지도 API데이터</p> <p>공공데이터</p>		<p>고객접촉채널</p> <p>모바일 앱</p> <p>웹사이트</p> <p>메타버스</p> <p>분양행사 참여(모텔하우스)</p> <p>온라인 SNS(인스타그램, 유튜브)</p>	
<p>비용소모 요소</p> <p>개발비(개발 비용(인건비))</p> <p>플랫폼 운영비(서버비용, 사무실, 인건비)</p>			<p>매출 발생 요소</p> <p>매물 등록 비용(개인)</p> <p>분양대행사(법인), 협찬</p> <p>웹사이트 비용(프리미엄 회원)</p>	

아이집사 SWOT 분석

<p>내부</p> <p><b>S : 강점</b></p> <ul style="list-style-type: none"> <li>• 나에게 맞는 집을 찾는 시간을 절약</li> <li>• 직접 가보지 않고도 쉽게 정보를 얻을 수 있다.</li> <li>• 정확한 데이터</li> <li>• 고객의 취향에 맞추어 적절한 데이터를 제공(맞춤형 데이터 ex, 넷플릭스)</li> <li>• 팀원의 부동산 전문 지식</li> </ul> <p><b>SO전략</b></p> <ul style="list-style-type: none"> <li>• 수요자입장을 고려한 서비스 특화 전략</li> <li>• 부동산 관심도 상승에 맞춘 시간 절약 서비스 제공</li> </ul> <p><b>ST전략</b></p> <ul style="list-style-type: none"> <li>• 공인중개사의 접속개 시간단축</li> <li>• 공인중개사 매물 홍보대행 서비스</li> <li>• 서비스 이용자 수수료 절감</li> <li>• 인공지능을 활용한 매칭 서비스를 공인중개사 매출 증대</li> </ul>	<p>외부</p> <p><b>O : 기회</b></p> <ul style="list-style-type: none"> <li>• MBC구체화 출시, 건축의 시간, 유튜브에서 부동산 관련 관심도가 오르고 있음</li> <li>• 부동산 가격상승, 코로나 사태로 자가 주택에 관심도 증가</li> <li>• 인공지능 기술발전과 사업 지원 정책</li> </ul> <p><b>T : 위협</b></p> <ul style="list-style-type: none"> <li>• 기존 부동산 IT업체들, 경쟁업체들이 선점 강박</li> <li>• 공인중개사 협회의 시위</li> </ul>
<p><b>W : 약점</b></p> <ul style="list-style-type: none"> <li>• 데이터 확보의 어려움 (DB확보)</li> <li>• 공인중개사 측의 부정적 반응</li> <li>• 창업 자금 부족</li> <li>• 개발자의 부족</li> </ul> <p><b>WO전략</b></p> <ul style="list-style-type: none"> <li>• 창업 자금 부족을 정부지원사업 참여로 해결</li> <li>• 기존 부동산 관련 IT업체와 협업을 통한 데이터 확보</li> <li>• 코로나로 인해 직접 가지 못하는 공인중개사에게 비대면 지원 서비스</li> <li>• 인공지능 시장 확대로 인한 외부 개발자 영입 (외주 개발 가능)</li> </ul> <p><b>WT전략</b></p> <ul style="list-style-type: none"> <li>• 지역 거점별 공인중개사와 협업 모델 구축을 통한 긍정적 참여 유도</li> <li>• 초기 데이터 확보를 위해 공인중개사 혜택 서비스(프리미엄 혜택 - 검색한도, 매물등록비免除)</li> <li>• 기존 부동산IT업체와 Consortium 협의</li> </ul>	

## 아이집사 AI기반 부동산 매물 추천 서비스 플랫폼 구성도



## 아이집사 서비스 AI 기술요소

<p><b>AutoML 솔루션 개발</b></p> <p>기계학습 등 인공지능 관련 모델 개발에 필요한 다양한 과정을 자동화하여 최적 성능의 모델 개발</p>	<p><b>RPA System 구성</b></p> <ul style="list-style-type: none"> <li>• <b>Robotic Process Automation!</b></li> <li>• 일반화되는 등 디지털 사무 환경이 실현되면서 기존 단순 업무로 로보틱 처리 자동화(RPA)가 대체</li> <li>• 부동산을 건 데이터처리, 지도위치 확인 등</li> </ul>	<p><b>Block Chain기반 서비스</b></p> <ul style="list-style-type: none"> <li>• <b>블록체인기반 부동산 서비스</b></li> <li>• <b>탈중앙화 신원증명(DID)</b> : 사용자 개인 정보를 자신의 단말기에 저장하고, 인증할 때 필요한 정보만 골라 제출할 수 있게끔 해주는 블록체인 기반의 전자 신원증명 기술</li> </ul>
<p><b>VR기반 서비스</b></p> <ul style="list-style-type: none"> <li>• VR기반 부동산매물 소개제작</li> <li>• VR기반 매물확인 서비스</li> <li>• VR기반 청약 모델하우스 서비스</li> </ul>	<p><b>AI기반 부동산 추천 서비스</b></p> <ul style="list-style-type: none"> <li>• ML 기반 모델개발</li> <li>• 자연어처리</li> <li>• 데이터 취합 및 수집</li> <li>• 데이터 전처리</li> <li>• Web/App연계 개발 및 테스트</li> </ul>	<p><b>Prop Tech기반 서비스</b></p> <ul style="list-style-type: none"> <li>• <b>프롭테크(Prop Tech)</b>는 부동산(property)과 기술(technology)을 결합하여</li> <li>• 정보 기술을 결합한 부동산 서비스 산업</li> <li>• <b>프롭테크</b> 비즈니스 영역은 크게 중개 및 임대, 부동산 관리, 프로젝트 개발, 투자 및 자금 조달</li> </ul>