

인공지능 프로젝트과정 이해와 산출물의 요소



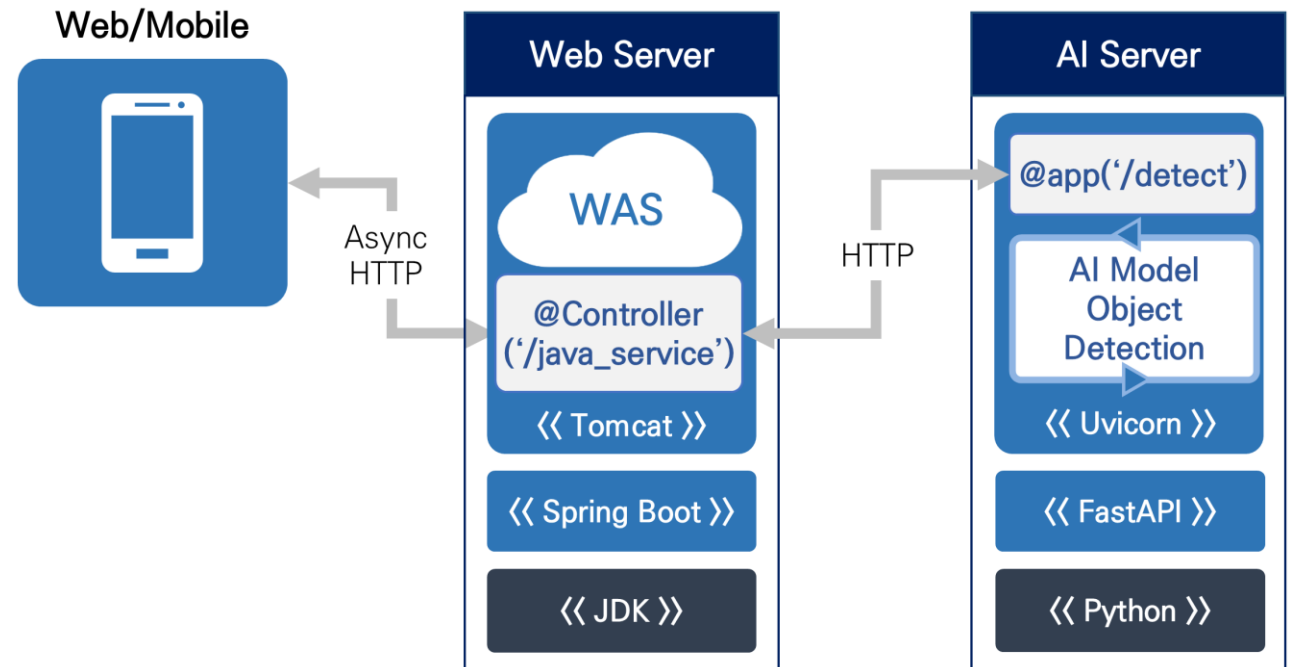
AI PROJECT
DOCUMENTS
SKILL, SERVICES



인공지능 프로젝트 유형

인공지능 프로젝트과정 이해와 산출물의 요소

- ✓ 모델을 직접 구현하고 기존 모델과 비교하는 유형
- ✓ 이미 구현된 모델을 인공지능 플랫폼에서 가져와 파인튜닝하는 유형
- ✓ OpenAPI를 사용하는 유형
- ✓ 웹으로 인공지능 서비스를 제공하는 유형
- ✓ 자바 웹과 파이썬 인공지능 플랫폼을 연동하는 유형



주제 선정

인공지능 프로젝트과정 이해와 산출물의 요소



프로젝트 주제

- ▶ 프로젝트명
- ▶ 프로젝트 개요
- ▶ 팀 이름 및 팀원 구성
- ▶ 데이터 출처
- ▶ 사용 기술 및 아키텍처



- 프로젝트 주제가 정해졌으면
 - 어떤 모델을 이용할지(찾아서 사용하거나, 아키텍처를 정의하거나) 구상해야 하고,
 - 프로젝트 개요를 작성하셔야 합니다.
 - 프로젝트 개요에는 아래 내용이 들어가도록 작성하세요.
 1. 주제 및 선정 배경(또는 기획의도)
 2. 프로젝트에 사용한 개발환경(장비 및 재료)과 기술 스택을 기록하세요. (개발환경은 버전이 기록되어야 합니다.)
 3. 프로젝트 구조를 구상했다면 도식화하세요.

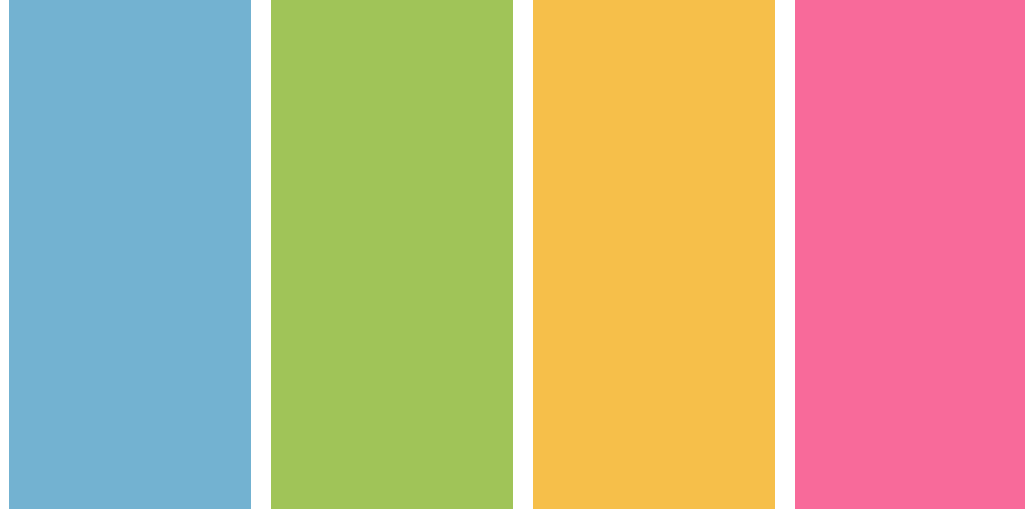
페이지 번호를 삽입하세요.
(위치는 어디가 되든지 상관 없습니다.)

작성 방법

인공지능 프로젝트과정 이해와 산출물의 요소

- ▶ 본 포트폴리오 양식은 프로젝트 팀 별로 각각 작성하여 제출
- ▶ 프로젝트 수행 과정 및 결과에 대해서는 제공된 목차 및 세부 항목별 작성요령을 참조하여 작성
- ▶ 프로젝트 특성에 따라 기본적인 구성을 유지한 상태에서 제공 양식을 보완하거나 추가하여 작성할 수 있음
- ▶ 본 슬라이드의 템플릿을 사용하는 것은 지양
- ▶ 작성 예시 및 작성요령 등은 모두 삭제 후 제출

💡 파일 -> 더보기 -> 옵션 -> 저장 -> '파일의 글꼴 포함'에 체크를 해주세요.
• Mac 사용자는 윈도우에서 프레젠테이션 전에 글꼴이 적용되는지 확인하세요.

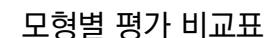
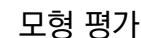
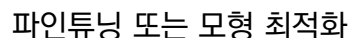
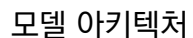
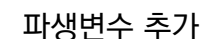
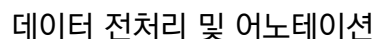
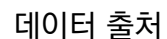
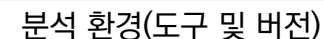
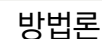


인공지능 프로젝트 발표 슬라이드 예시

인공지능 프로젝트과정 이해와 산출물의 요소



프로젝트 개요(도메인 이해가 필요)

6



표지

인공지능 프로젝트과정 이해와 산출물의 요소

- 프로젝트 주제(프로젝트 명)
- 팀명
- 소속 및 이름

PPT의 크기는 Wide(권장) 또는 A4가로



목차

인공지능 프로젝트과정 이해와 산출물의 요소

- 장 제목(대 주제)과 절 제목(중 주제)
- 목차 소개 시 대 주제만 언급하고 지나감
- 발표시 예: 제안 개요부터 사업수행전략, 사업화 방안 그리고 사업관리방안 순으로 설명하겠습니다.





설명 슬라이드의 내용

인공지능 프로젝트과정 이해와 산출물의 요소

- 장 제목과 절 제목 그리고 슬라이드의 제목
- 슬라이드를 설명하는 요약텍스트 및 상세 텍스트
- 관련 아이콘 및 이미지
- 상세 설명 및 도표 등 포함





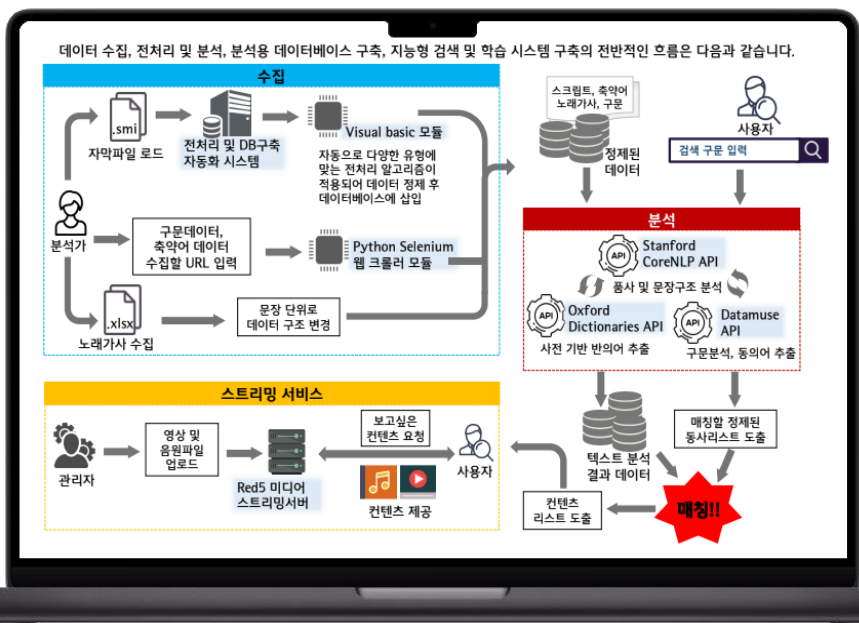
프로젝트 흐름도 또는 시스템 아키텍처

인공지능 프로젝트과정 이해와 산출물의 요소

- 프로젝트의 전체 흐름을 한눈에 파악할 수 있도록 프로젝트 흐름도를 작성
- 직관적이고 자세하게 작성
- 분석의 절차 및 시스템의 구성도를 포함할 수 있음



무엇보다 중요한 것은 전체 프로젝트 흐름도 또는 구성도입니다.



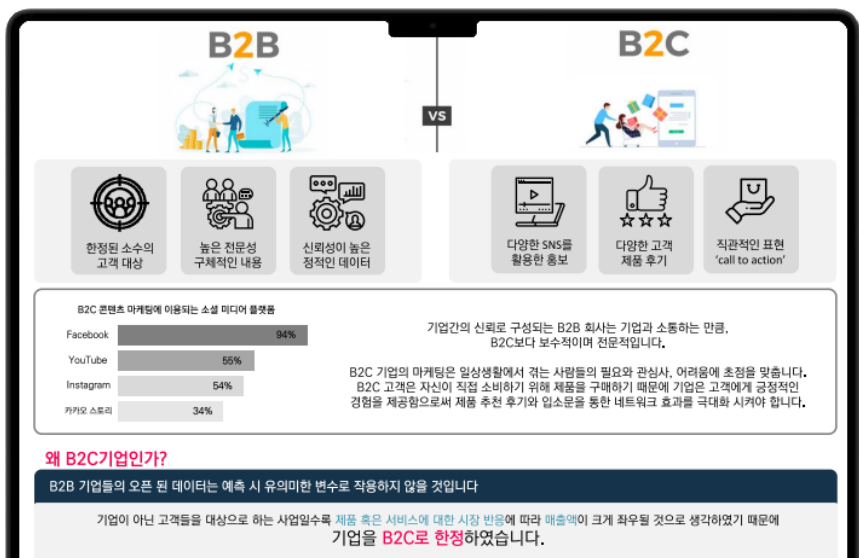


프로젝트 개요

인공지능 프로젝트과정 이해와 산출물의 요소

- 프로젝트의 목적 및 개요에 대해 기술
- 프로젝트 주제 및 선정 배경(기획의도 등) 및 기대효과
- 프로젝트 개요(프로젝트 구현 내용, 컨셉, 훈련내용과의 관련성 등)
- 활용 장비 및 재료(개발 환경 등)
- 프로젝트 구조

B2B, B2C, As-is To-be 등이 다음 슬라이드에 포함될 수 있음





프로젝트 팀 구성 및 역할

인공지능 프로젝트과정 이해와 산출물의 요소

- [프로젝트 팀 구성 및 역할]은 프로젝트를 기본 단위로 작성하며 팀원의 수에 따라 칸을 추가/삭제할 수 있다.
- 담당 업무 : 팀원 별로 해당 프로젝트를 진행하면서 주도적으로 참여한 부분을 중심으로 작성

팀원이 1명일 경우도 본인의 역할을 모두 기록

훈련생	역할	담당 업무
홍길동	팀장 및 팀원	▶ 데이터 정제 및 정규화
		▶ 모바일 서비스 테스트
		▶ 모바일 플랫폼 구현
		▶ 외부 데이터 수집
		▶ 서비스 시스템 설계
		▶ 텍스트 마이닝

훈련생	역할	담당 업무
김○○	팀장	▶ 데이터 정제 및 정규화 ▶ 모바일 서비스 테스트
박○○	팀원	▶ 모바일 플랫폼 구현 ▶ 외부 데이터 수집
정○○	팀원	▶ 서비스 시스템 설계 ▶ 텍스트 마이닝



프로젝트 수행 절차 및 방법

인공지능 프로젝트과정 이해와 산출물의 요소

- [프로젝트 수행 절차 및 방법]은 프로젝트의 사전 기획과 프로젝트 수행 및 완료 과정으로 나누어서 작성.
- 프로젝트 수행 절차를 도식화하여 제시하거나, 더 효과적으로 전달하는 방법 등이 있다면 수정하여 작성 가능
- 기획 단계에서 도출된 주제와 아이디어를 기반으로 실제 프로젝트를 수행한 세부적인 기간과 활동 내용 작성

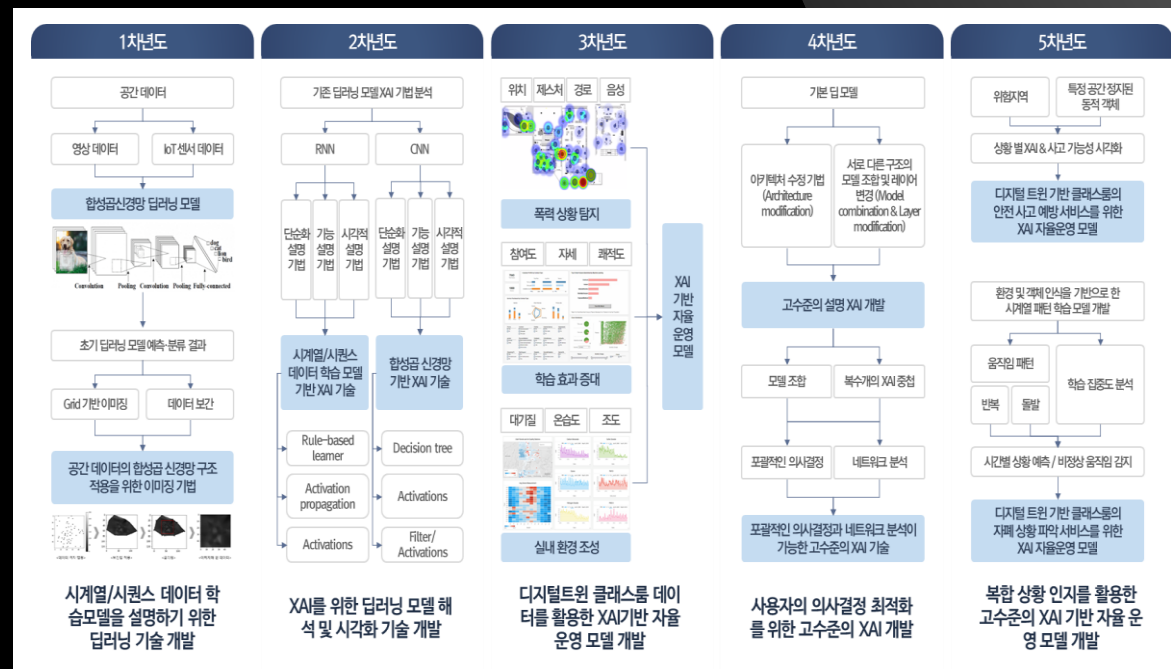
프로젝트 목적/목표와 함께 한 장의 장표로 가능

프로젝트 목적 : 교육 내용 전반을 활용한 예측 모델 개발 및 웹 서비스 프로토타입 구현

프로젝트 목표 : 섭취 식품 및 건강 설문 응답 데이터에서 질환 보유자의 패턴을 학습하고, 이를 근거로 일반 사용자의 질환 가능성 예측 결과를 제공할 수 있다.

프로젝트 범위 : 데이터 크롤링 / DB구축 & SQL처리 / 머신러닝 & 딥러닝 / 질환 예측 / 웹 구현

	11												12																													
	3주차				2주차				3주차				4주차				5주차				6주차																					
	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	1	2	3	4	5	6	7	8	9	10	11	12	13		
	월	화	수	목	금	토	일	월	화	수	목	금	토	일	월	화	수	목	금	토	일	월	화	수	목	금	토	일	월	화	수	목	금	토	일	월	화	수	목	금	토	일
0	교료직보 주제 선정																																									
1	레이더 크롤링 / 수집																																									
2	EDA (탐색적 자료분석)																																									
3	레이더 전처리																																									
4	모델 생성 / 파라미터 최적화																																									
5	파생변수 탐색 및 추가																																									
6	DB 구축 / SQL 처리																																									
7	Django 웹 구현																																									
8	발표자료 작성 / 발표 준비																																									





방법론

인공지능 프로젝트과정 이해와 산출물의 요소

- 프로젝트 방법론을 정의 함(선택사항)
- SEMMA 방법론만 있는 것은 아님
- 데이터 이해 단계가 포함되어야 한다면 CRISP-DM 방법론 사용
- 응용소프트웨어 개발방법론이 추가될 수 있음





분석 환경 및 도구 (기술 스택)

인공지능 프로젝트과정 이해와 산출물의 요소

- 분석 환경 또는 개발환경에 사용한 도구들 정리
- 이름과 버전을 명시해야 함
- 시스템 및 H/W
- 프로그래밍 언어 및 데이터베이스
- 개발도구, 라이브러리, 프레임워크





데이터

인공지능 프로젝트과정 이해와 산출물의 요소

- 데이터의 출처 및 크기 등을 표기
- 어떤 사이트에서 어떤 데이터를 어떤 방법으로 어떤 도구를 이용해서 수집할 것인가?
- 자연어처리
 - CSV 파일 형식, 문장 형식, 어노테이션이 있는 형식
 - 크기(MB), 행의 수 -> 임베딩한 후의 배열의 크기
- 이미지
 - 컬러영상, 이미지 개수, 크기(MB)
 - 어노테이션 형식(xml, json, txt) - 예시 캡처 포함

	MAIN 질병관리본부 KCDC	SUB h-well 국민건강보험	Food 식품안전나라	보건복지부
정의	국민건강영양조사 원시 데이터	건강검진 데이터	식품 영양소 데이터	한국인 영양소 섭취 기준
내용	건강설문, 식품섭취빈도 등 패널조사 응답 결과	공단 의무검진 항목별 측정 결과	식품별 영양소 함유량	연령 및 성별에 따른 영양소 섭취 기준
시점	2007년 ~ 2017년	2002년 ~ 2017년	2009년 ~ 2016년	2015년 11월
형태	SAS File	Excel File	Web	PDF File
신뢰성	국가승인통계 (승인번호 제117002호)	국가기관 발행 데이터 (최초등록-2015/07/24)	공공기관 연구 자료 및 국가기관 구축 DB	국가기관 발행 데이터 (등록번호 11-1352000-001537-14)

“ 질환 예측을 위한 분류분석용 데이터로 활용 (개인정보 및 질환진단 데이터)” ”	“ 식품 섭취에 따른 영양소 섭취량 산출 데이터로 활용” ”
---	--



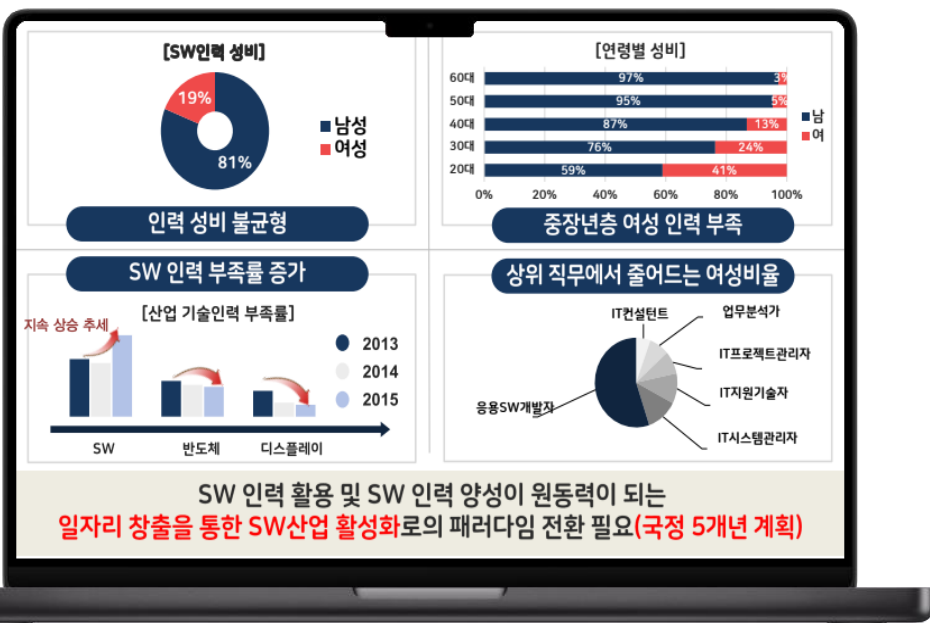


데이터 탐색

인공지능 프로젝트과정 이해와 산출물의 요소

- 탐색적 데이터 분석 수행
- EDA 결과를 시각화 할 필요가 있음

시각화 결과를 어떻게 정리할 것인가를 미리 생각해야 함.





데이터 전처리

인공지능 프로젝트과정 이해와 산출물의 요소

- 결측치, 이상치 탐색 및 제거
- 표준화 또는 정규화
- 레이블 인코딩, 원-핫 인코딩
- 전처리 대상 열의 선택 기준 정의
- 다른 값을 대체할 경우 이유와 타당성 명시

ML/DL 데이터에 따라 다양한 전처리 방법이 사용됨

01 원문 데이터

※자동차 주행중에 소리를 찍을수없어 최대한 비슷한 소리로 올려봅니다.
<http://cafe.naver.com/morningjiang/969137>
 Datachannel : café

02 문서 정제

※자동차 주행중에 소리를 찍을수없어 최대한 비슷한 소리로 올려봅니다.
<http://cafe.naver.com/morningjiang/969137>

03 자연어 처리

자동차[명사] 주행[명사] 중[명사] 예[조사] 소리[명사] 들[조사] 찍을[동사]
 수[명사] 없어[동사] 최대한[부사]
 비슷한[형용사] 소리[명사] 로[조사] 올려봅니다[동사].

04 키워드 및 연관어 추출

자동차: 자동차 카테고리
 비슷한: 형용사 중립 Control
 올려보다: 동사 중립 Control

1. total_parking_capacity_in_site (아파트 단지 내 총 주차 대수)

2. tallest / lowest_building_in_sites (아파트 단지 내 최고/최저 층)

3. heat_type / heat_fuel (난방 타입/난방 연료)

4. room_count / bathroom_count (방 수/화장실 수)

5. front_door_structure (앞문 앞 구조)

6. subways.csv - address_by_law (법정동)

Case 1

Case 2

Case 3

데이터 요약 통계량 탐색 후 데이터 전처리를 어떻게 할 것인가?

데이터 분석을 그제로 다시 돌려서 변경할 항목을 다시 확인

데이터의 중요도를 높이고, 데이터의 과묵을 피하기는 시간

변수명 이전 (2007 - 2011)

변수명 이후 (2012 - 2014)

데이터 사용



프로젝트 수행 결과

인공지능 프로젝트과정 이해와 산출물의 요소



[프로젝트 수행 결과]는 프로젝트 결과물이 도출된 과정을 세부적으로 기록

- 예시는 하나의 사례로 간단하게 제시한 것이므로 프로젝트의 성격에 따라 보다 자세하게 기록하며, 결과를 서술하는 과정에서는 활용된 기술(구현 방법), 핵심기능, 검증 결과* 등을 상세히 기재한다.

* 예. 인공지능 객체탐지일 경우 IoU50 mAP 0.9, 빅데이터 직종의 경우 정확도, f1-score 등

- 프로젝트의 결과는 그 과정이 잘 드러날 수 있도록 가공 과정부터 활용까지 전체적인 프로세스를 확인할 수 있도록 단계별로 작성

* 첨부 자료 예시: 결과물 사진, 시연 동영상 등 프로젝트의 우수성이 드러날 수 있는 자료



프로젝트 수행 결과 - 머신러닝 기법 또는 딥러닝 모델

인공지능 프로젝트과정 이해와 산출물의 요소

- 머신러닝/딥러닝 기법 및 기본 이론에 대한 설명 추가
- 모델을 평가한 결과 기록
- 딥러닝의 경우 딥러닝 모델을 도식화 함

유사 모델을 비교 분석하는 것도 필요함

원-핫 인코딩 단어를 학습시키기 위해 벡터로 변환하는 작업 중에서 가장 간단한 방법으로, 0 또는 1로 변환합니다.

BUT

- 단어간의 유사도를 파악할 수 없습니다.
- 모든 벡터가 수직이므로 cosine similarity = 0
- 3개의 단어는 3차원의 벡터가 되므로 차원이 너무 커집니다.

임베딩 원-핫 인코딩의 단점을 보완한 것입니다.

- 원-핫 인코딩보다 저차원의 벡터를 생성합니다.

Unique word	Encoding	Embedding
Kine	[1, 0, 0, 0]	[1, 2]

모델을 직접 구현하지 않았더라도 모델의 구조를 표현하고 설명할 수 있어야 합니다.

VS

주변에 있는 단어로 중간에 있는 단어를 예측합니다

중간에 있는 단어를 가지고 중간에 있는 단어를 예측합니다.

모델을 선정한 이유

- 모델의 구조를 표현해야 합니다.
- 구현한 소스코드 또는 summary() 결과를 활용하세요.
- 평가 방법에 대한 설명이 추가되어야 합니다.

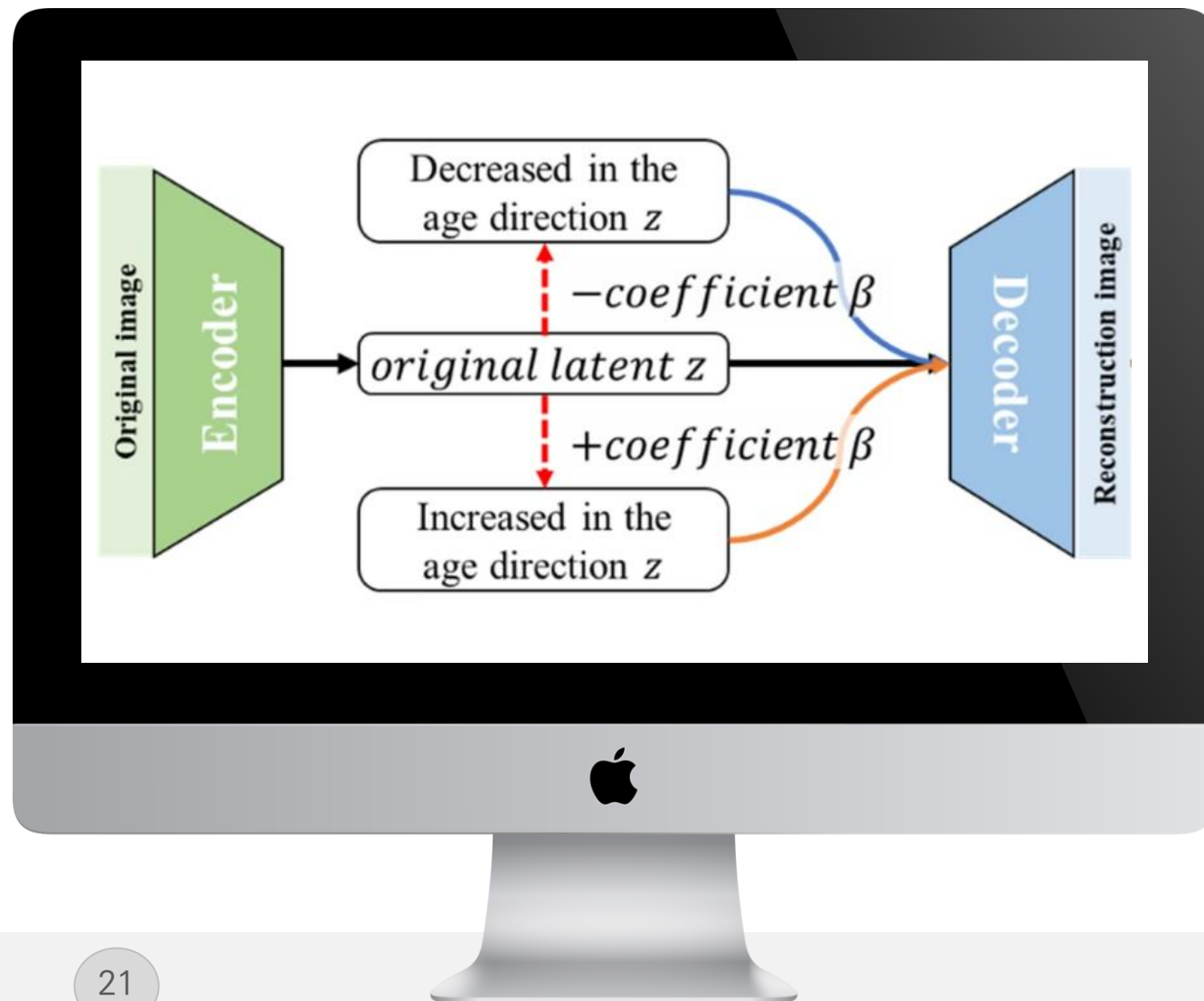
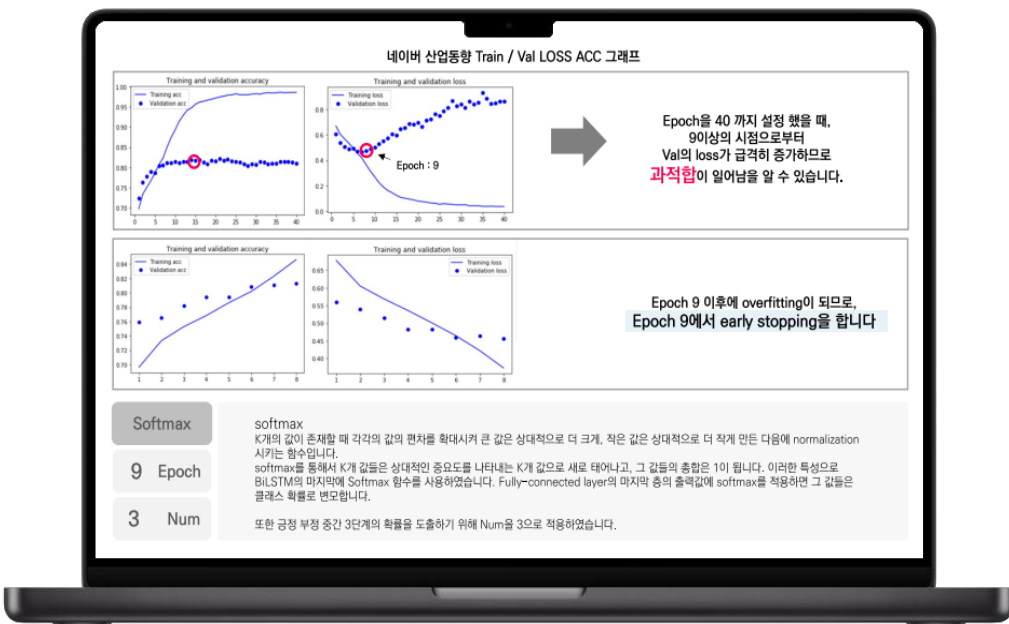


프로젝트 수행 결과 - 파인 튜닝 또는 모형 최적화

인공지능 프로젝트과정 이해와 산출물의 요소

- 모델의 평가 방법 및 평가 점수 개선을 위한 노력이 포함되어야 함
- 모델의 수정 또는 하이퍼파라미터 조정 등 작업이 필요
- 과적합 여부 검사
- 모델을 비교할 수 있는 엑셀파일을 만들고 기록

최적화 전/후의 모형을 비교



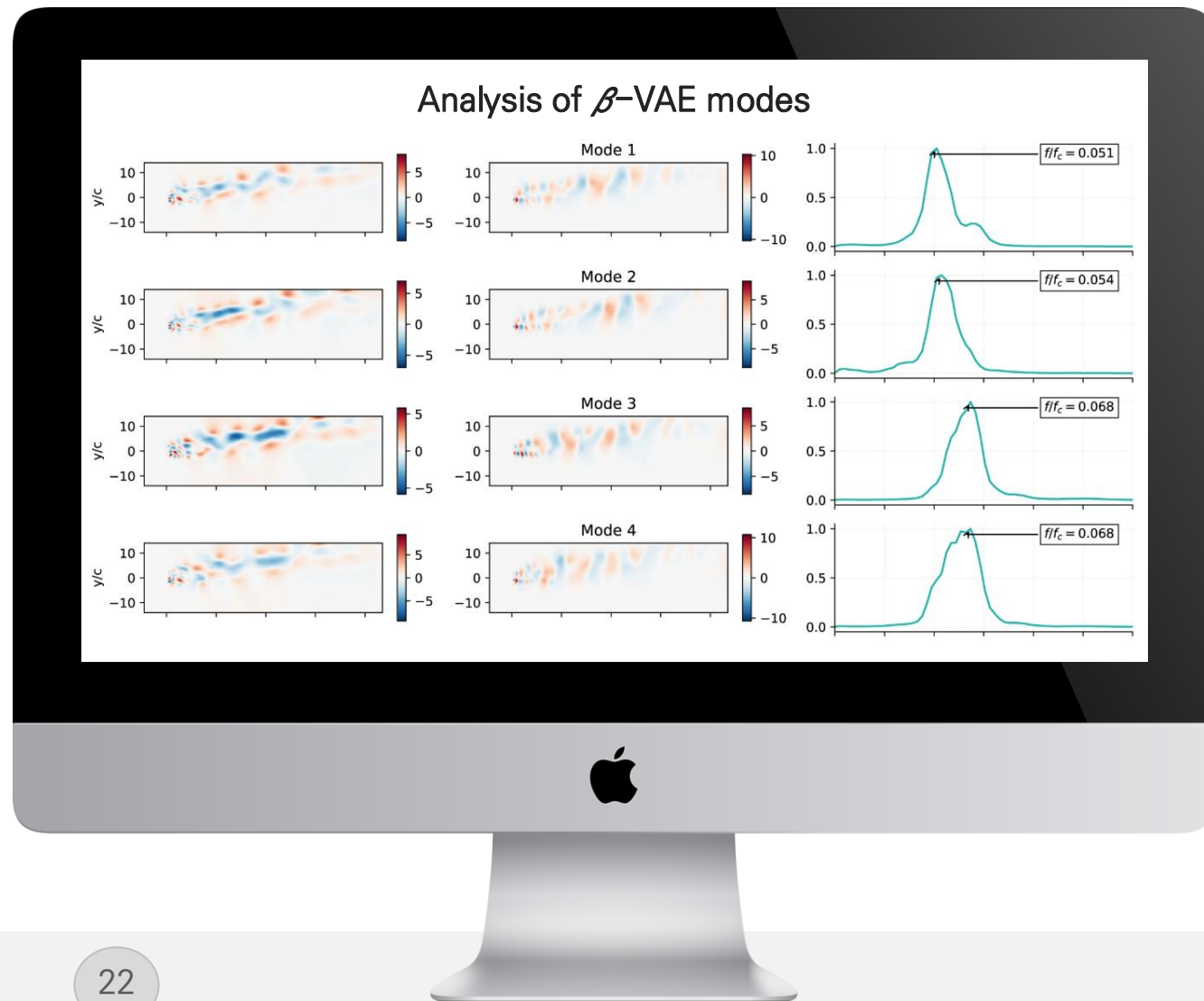
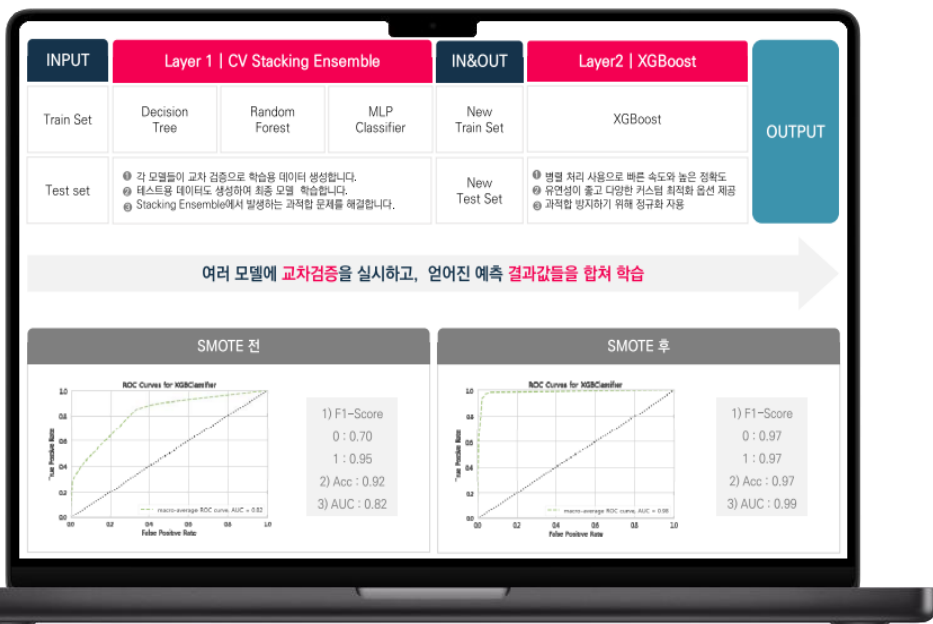


프로젝트 수행 결과 - 모형 평가 비교

인공지능 프로젝트과정 이해와 산출물의 요소

- 모형을 평가하는 방법을 표시
- 평가방법을 선택한 근거와 이유 제시
- 교차검증 실시

교차검증을 통한 모형 평가가 필요할 수 있음





프로젝트 수행 결과 - 모형별 평가표

인공지능 프로젝트과정 이해와 산출물의 요소

- 데이터의 크기와 알고리즘, 평가 점수를 기록해야 함
- 전처리 방법에 따른 평가점수
- 파생변수들 추가에 따른 평가점수
- 분석/예측 모형에 따른 평가점수

비교표를 제시해야 함

SMOTE 전				SMOTE 후				결론	
Decision Tree				Decision Tree				SMOTE 전 F1-Score, Acc RF (Stacking Ensemble AUC RF) Stacking Ensemble → Stacking Ensemble 채택	
	F1-Score	Acc	AUC		F1-Score	Acc	AUC		
0	0.59	0.85	0.78	0	0.90	0.90	0.90	SMOTE 후 F1-Score, Acc RF (Stacking Ensemble AUC RF) Stacking Ensemble → Stacking Ensemble 채택	
1	0.91	0.85	0.78	1	0.90	0.90	0.90		
Random Forest				Random Forest				최종 채택 Model 서비스 목적 : 성공(1) 예측 1에 대한 F1-Score를 SMOTE 전, 후로 비교했을 때 큰 차이가 없음 Data의 변형 위험을 감소하며 SMOTE를 사용하지 않고 SMOTE 전 Model 채택	
	F1-Score	Acc	AUC		F1-Score	Acc	AUC		
0	0.54	0.85	0.91	0	0.96	0.96	0.99		
1	0.91	0.85	0.91	1	0.96	0.96	0.99		
MLPClassifier				MLPClassifier					
	F1-Score	Acc	AUC		F1-Score	Acc	AUC		
0	0.68	0.90	0.50	0	0.92	0.92	0.50		
1	0.94	0.90	0.50	1	0.92	0.92	0.50		
CV Stacking Ensemble(XGBoost)				CV Stacking Ensemble(XGBoost)					
	F1-Score	Acc	AUC		F1-Score	Acc	AUC		
0	0.70	0.92	0.82	0	0.97	0.97	0.98		
1	0.95	0.92	0.82	1	0.97	0.97	0.98		

Model Results													
모 델 링	Baseline	Library	Validation set	Without New_whale	Real-time Augmentation (Train)	Real-time Augmentation (Test)	Bounding -Box	Image padding + margin b-box	흑백 (grey)	Ensemble	Image Size (H/W)	batch_size	Score
#0	Sample submission												
#1	CNN	Keras	x	X	X	X	X	X	X	X	100x100	64	0.277
#2	ResNet 18	Fast.ai	5개 미만증식 (복사)	X	O	O	X	X	X	X	384x384	64	0.290
#3	ResNet 18	Fast.ai		O	O	O	X	X	X	X	384x384	64	0.565
#4	ResNet 34	Fast.ai		X	O	O	X	X	X	X	384x384	64	0.562
#5	ResNet 34	Fast.ai		O	O	O	X	X	X	X	384x384	64	0.588
#6	ResNet 50	Fast.ai		X	O	O	X	X	X	X	382x382	*32	0.562
#7	ResNet 50	Fast.ai	사전증식 (랜덤증식)	O	O	O	X	X	X	X	382x382	*32	0.577
#8	ResNet 34	Fast.ai		X	O	O	O	X	X	X	384x384	64	Best score 0.614
#9	ResNet 34	Fast.ai		O	O	O	O	X	X	X	384x384	64	0.590
#10	ResNet 34	Fast.ai		O	O	O	O	X	X	X	384x384	64	0.588
#11	ResNet 34	Fast.ai		O	O	O	X	O	X	X	384x384	64	0.562
#12	ResNet 34	Fast.ai	Model Ensemble (3단계)	O	O	O	X	O	O	X	384x384	64	0.573
#???	ResNet 34	Fast.ai		X	O	O	O	X	O	X	384x384	64	
#???	ResNet 34	Fast.ai											



프로젝트 수행 결과 - 시연

인공지능 프로젝트과정 이해와 산출물의 요소

- 데이터의 크기와 알고리즘, 평가 점수를 기록해야 함
- 전처리 방법에 따른 평가점수
- 파생변수들 추가에 따른 평가점수
- 분석/예측 모형에 따른 평가점수

※ 별도 파일로 첨부 가능



- 제출할 발표 영상은 5분 분량으로 만들어주세요.
- 시연 영상을 만들 때 자막을 포함하면 영상을 보는 사람이 더 이해하기 쉽습니다.



자체 평가 의견 또는 프로젝트 후기

인공지능 프로젝트과정 이해와 산출물의 요소

- ▶ [자체 평가 의견]은 프로젝트 결과물에 대한 프로젝트 기획 의도와의 부합 정도 및 실무 활용 가능 정도, 달성도, 완성도 등 자체적인 평가 의견과 느낀 점을 작성한다.
 - 개인 또는 우리 팀이 잘한 부분과 아쉬운 점을 작성한다.
 - * 예. 모델 평가 결과, 정확도가 00.00%로 정확도 향상을 위해 모델 추후 개선 필요
 - 프로젝트를 수행하면서 느낀 점이나 경험한 성과에 대하여 기재할 수 있으며, 경력 계획 등과 연관시켜 팀별 공통 의견 또는 개인 의견을 자유롭게 작성한다.