



4차산업 · 지능 정보화 시대의 필수역량
데이터 수집 · 처리 · 분석을 위한
“2020 K-디지털 파이썬 기반 응용AI 개발자 양성과정”

미니 프로젝트

(소비자 분석)

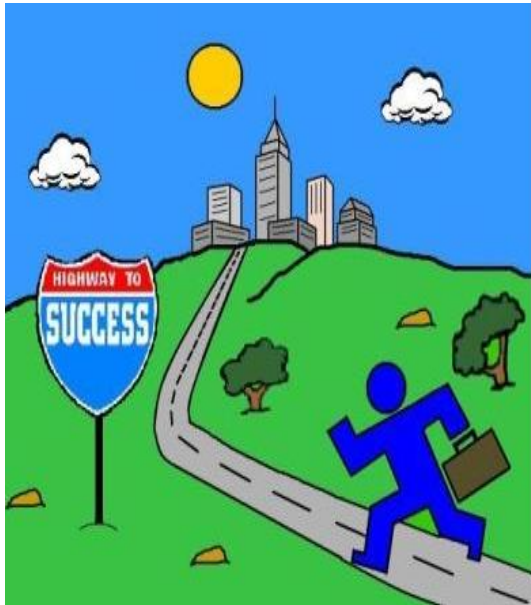
2021년 1월

우준식 (주) 시엠아이코리아 센터장

CISA / PMP / DAsP/ 정보시스템감리원

작업상담사 / 경영지도사

Hp) 010-5351-6791 / jswoo100@empas.co.kr



21.01.25 월요일
(미니 프로젝트 과정)

- ✓ Zoom으로 수업 입장
- ✓ 09:30~13:20 오전 프로젝트 진행
- ✓ 13:30~14:30 점심
- ✓ 14:30~18:20 오후 프로젝트 진행

- ✓ 자리에 있을때는 화면 켜놓기
- ✓ 시작전
 - 해당 날짜에 주간/일일 계획 점검
- ✓ 팀 or 개인 프로젝트
 - 1인 혹은 2인 프로젝트
- ✓ 질문은 수시

✓월요일 -> 계획 수립 및 발표

✓수요일 -> 진행 사항 발표

✓목요일 -> 진행 사항 발표, 내부 발표

프로젝트 선정

✓금요일 -> 프로젝트 완료 보고 및 발표

✓ 할 수 있는 주제 vs 하고 싶은 주제

✓ 과정 vs 결과

✓ 일일 진행 현황 정리

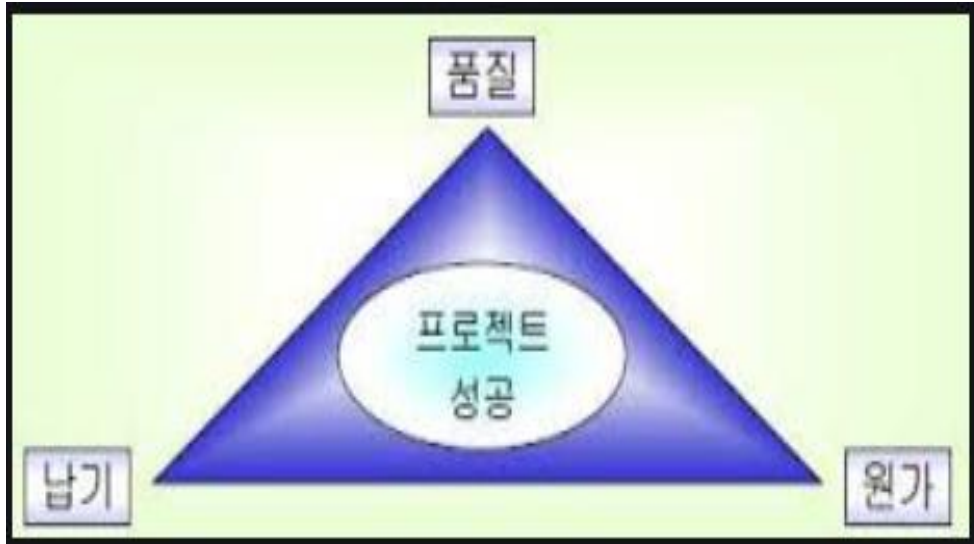
단계	단계 설명	수행 업무
분석 기획	해당 도메인의 문제점을 식별하고 이를 바탕으로 분석 프로젝트의 범위를 확정하는 단계	<ul style="list-style-type: none"> 프로젝트 정의 및 범위 확정 프로젝트 계획 수립 등 <ul style="list-style-type: none"> 문제 정의 <ul style="list-style-type: none"> 주제 발굴 종속변수 vs 독립변수
데이터 준비	분석 기획 단계의 산출물을 토대로 관련 데이터를 조사하고 준비하는 단계	<ul style="list-style-type: none"> 데이터 현황 조사 필요 데이터 정의 데이터 수집 데이터 검증 <ul style="list-style-type: none"> 데이터 수집 <ul style="list-style-type: none"> 내부·외부 데이터 데이터 유형별 데이터 정제 (이상치, 결측치 등)
데이터 분석 및 검증	데이터 분석 프로세스를 진행하고 그 결과물을 검증하는 단계	<ul style="list-style-type: none"> 데이터 준비 및 탐색 데이터 전처리 모델링 모델 평가 및 검증 <ul style="list-style-type: none"> 데이터 탐색 <ul style="list-style-type: none"> 기술적 분석·탐색적(EDA) 분석 시각화 (그래프) 상관관계 분석 분석 모델링 <ul style="list-style-type: none"> 알고리즘 선택 (분류·예측·군집화) 알고리즘 평가
모델 적용	데이터 분석 프로세스의 결과물은 모델을 적용하는 단계	<ul style="list-style-type: none"> 서비스 설계 및 구현 서비스에 분석 모델 적용 서비스 및 시스템 테스트
배포	모델을 적용한 시스템을 실제 서비스가 가능하도록 배포하는 단계	<ul style="list-style-type: none"> 개발한 서비스의 서버 배포 배포한 애플리케이션 및 서비스 운영

데이터 분석 과정



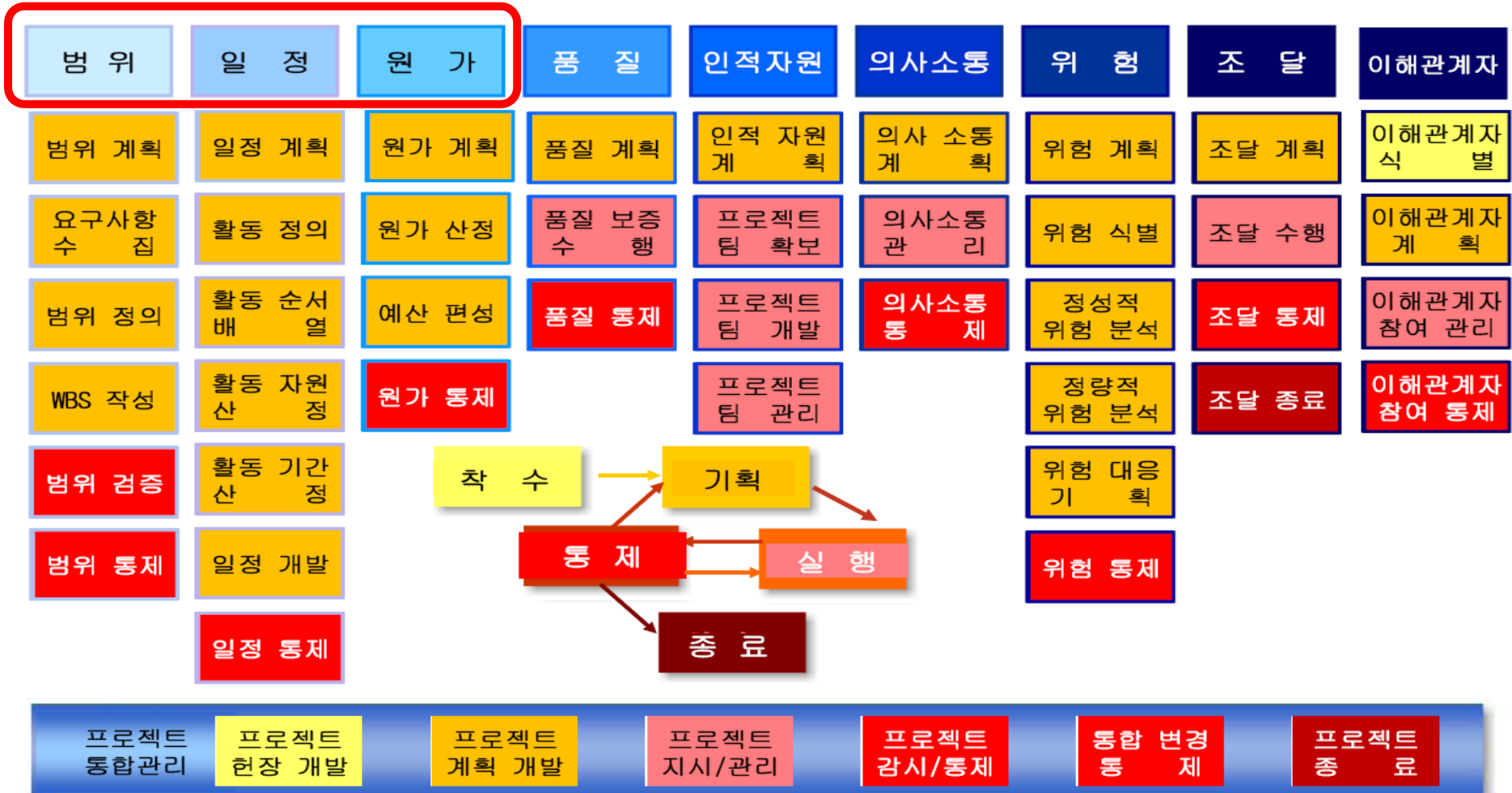
- | | | | | |
|--|--|---|--|--|
| <ul style="list-style-type: none"> ▪ 관심있는 주제 발굴 ▪ 종속변수(타겟) 설정 <ul style="list-style-type: none"> - 연속형 vs 범주형 ▪ 독립변수(피처) 리스팅 <ul style="list-style-type: none"> - 종속변수를 잘 설명해줄 것으로 예상되는 변수들 나열 (ideation) | <ul style="list-style-type: none"> ▪ 데이터 수집 <ul style="list-style-type: none"> - 보유 데이터 - 외부 데이터 구매 - 웹 크롤링 등 ▪ 데이터 정제 <ul style="list-style-type: none"> - 중복 데이터 제거 - 결측값(NA) 처리 등 ▪ 파생변수 생성 <ul style="list-style-type: none"> - 스코어링(grouping) 등 | <ul style="list-style-type: none"> ▪ 데이터 시각화 (EDA) <ul style="list-style-type: none"> - 산점도, 선/막대 그래프 - 히스토그램, 상자그림 등 ▪ 상관관계 분석 <ul style="list-style-type: none"> - 연속형(y) & 연속형(x) - 범주형(y) & 연속형(x) - 범주형(y) & 범주형(x) | <ul style="list-style-type: none"> ▪ 알고리즘 선택 <ul style="list-style-type: none"> - 분류모형 - 예측모형 ▪ 알고리즘 평가 <ul style="list-style-type: none"> - Hold-out vs Cross-validation - MSE(예측모형) - ROC(분류모형) 등 | <ul style="list-style-type: none"> ▪ 인사이트 발굴 및 정리 <ul style="list-style-type: none"> - 커뮤니케이션 - 데이터 시각화 (PT) - 프레젠테이션 ▪ 실행 및 모니터링 <ul style="list-style-type: none"> - 인사이트 기반 액션 플랜 - 현업에 적용 - 결과 모니터링 - 새로운 문제 정의 |
|--|--|---|--|--|

(나성호 연구원 자료)



- ☐ 프로젝트란
 - . 일정한 기간 안에
 - . 일정한 목적을
 - . 달성하기 위해 수행하는 업무의 묶음
- ☐ 프로젝트는 계획된 시간 안에 목표로 하는 범위를 완수해야 하는 작업
 - . 시작, 종료일이 정해져 있음
 - . 목적 혹은 수행할 작업의 범위가 명확히 명시 되어 있음.
 - . 예산이 미리 책정되어 있음.
 - . 프로젝트가 종료되면 해산하는 임시 조직이 수행
 - . 해결할 업무가 있음.
 - . 일련의 예상되는 효과
 - . 생산품, 결과, 산출물이 정해져 있음
 - . 역할 담당된 인력들이 업무를 수행

PMBOK의 프로젝트 관리 프로세스 - 5th Edition



◎. 프로젝트 진행하며 집중해야 할것

- ✓ 데이터 분석 프로세스 이해 하기
- ✓ 데이터 가공 기술 익히기
 - Python(함수/알고리즘), Pandas, Numpy, matplotlib, seaborn
 - 사용되는 라이브러리 및 Domain 데이터 처리(텍스트, 시계열, 이미지, 음성, 센서등)
 - jupyter notebook
 - vscode / pycharm / Spyder
- ✓ 데이터 파이프라인 기술 익히기
 - 수집 - 저장·관리 - 분석 - 시각화
 - 웹-DB(sql)/OS(linux, cloud, windows)-통계/알고리즘(머신러닝, 딥러닝)-이론 & Tool
- ✓ 문서화 기술
 - github 등 (source 버전 관리, 작업내용 기록 남기기)
- ✓ 에러 처리 및 문제발생 대응 능력 (Document 기반 해결 능력)

◎. 인공지능 서비스 종류

<표 1-2> 인공지능 서비스 종류

내용	내용
인지컴퓨팅	인간과 같이 정보를 습득하고 의사결정을 할 수 있는 모델을 모의실험하여 의사 결정을 할 수 있는 서비스
기계학습	프로그래밍된 논리 또는 정형화된 프로그램 규칙으로 생성된 데이터를 기반으로 학습하는 수학적 알고리즘 서비스
딥 러닝	인간 및 신경 네트워크를 모델링하고 새로운 데이터 세트를 예측하는 서비스
자연어 처리	인간의 언어를 배워서 인간처럼 사용할 수 있는 서비스
데이터 마이닝	의사결정을 위하여 대량의 데이터에서 내재되어 있는 상관관계를 발견하여 유용한 정보를 추출하는 서비스
음성인식	인간의 음성을 인식하여 문자로 변환하는 서비스
패턴인식	모양, 문자, 소리등을 인식하는 서비스
전자 눈	실세계의 사물을 인식하여 처리할 수 있는 전자 눈 서비스
이미지 패턴 매칭	어떤 이미지가 무슨 의미를 가지고 있는지 식별할 수 있는 서비스
체감형 현실 기술	컴퓨터에서 특정 환경이나 상황을 만들어 실제 환경과 상호 작용하는 것처럼 보이게 하는 휴먼-컴퓨터 인터페이스 서비스

◎. 인공지능 서비스 적용 사례



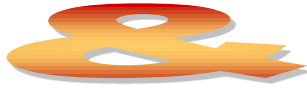
출처: 정보통신기술진흥센터. 국내외 AI활용 현황과 공공 적용(2018년). 20쪽
[그림 1-2] 국내 주요 산업별 인공지능 서비스

◎. 인공지능 서비스 구성 사례

인공지능 기기	인공지능 서비스 모델		인공지능 서비스	
로봇	인공지능 소프트웨어	챗봇, 음성인식, 영상인식, 통번역, 음성비서	전문가 서비스	법률, 금융, 의료 등
자동차	인공지능 플랫폼	기계학습, 딥러닝, 자연어처리, 데이터 처리	엔터테인먼트	음악, 영화, 게임 등
가전제품	인공지능 프레임워크	TensorFlow, Apache MXNet, Keras, Caffe, PyTorch	커머스	O2O, 쇼핑, 광고 등
IoT	인공지능 인프라	CPU, GPU, IoT, Cloud, Big Data, Mobile	IoT 서비스	스마트 카, 스마트 홈 등
모바일				
웨어러블 기기				

[그림 1-3] 인공지능 서비스의 구성 사례





우준식 [주]시엠아이코리아 아카데미 센터장

직업상담사 / 경영지도사(인사관리) / 인적자원개발사

PMP / CISA / 재난관리사 / 정보시스템감리원

“IT.Career · HRD 전문가”

연락처 : [hp] 010-5351-6791, jswoo100@empas.com

블로그 : <http://blog.naver.com/jswoo100>

SNS : [twitter ID]->@goodjob21

[facebook]->jswoo100@empas.com