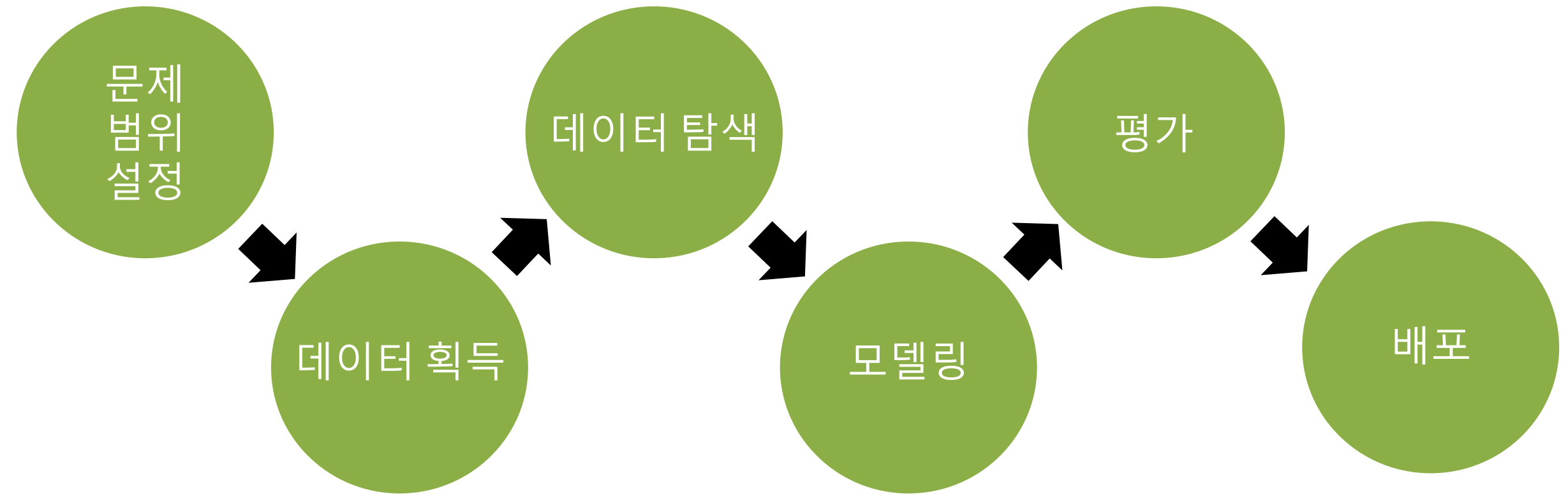
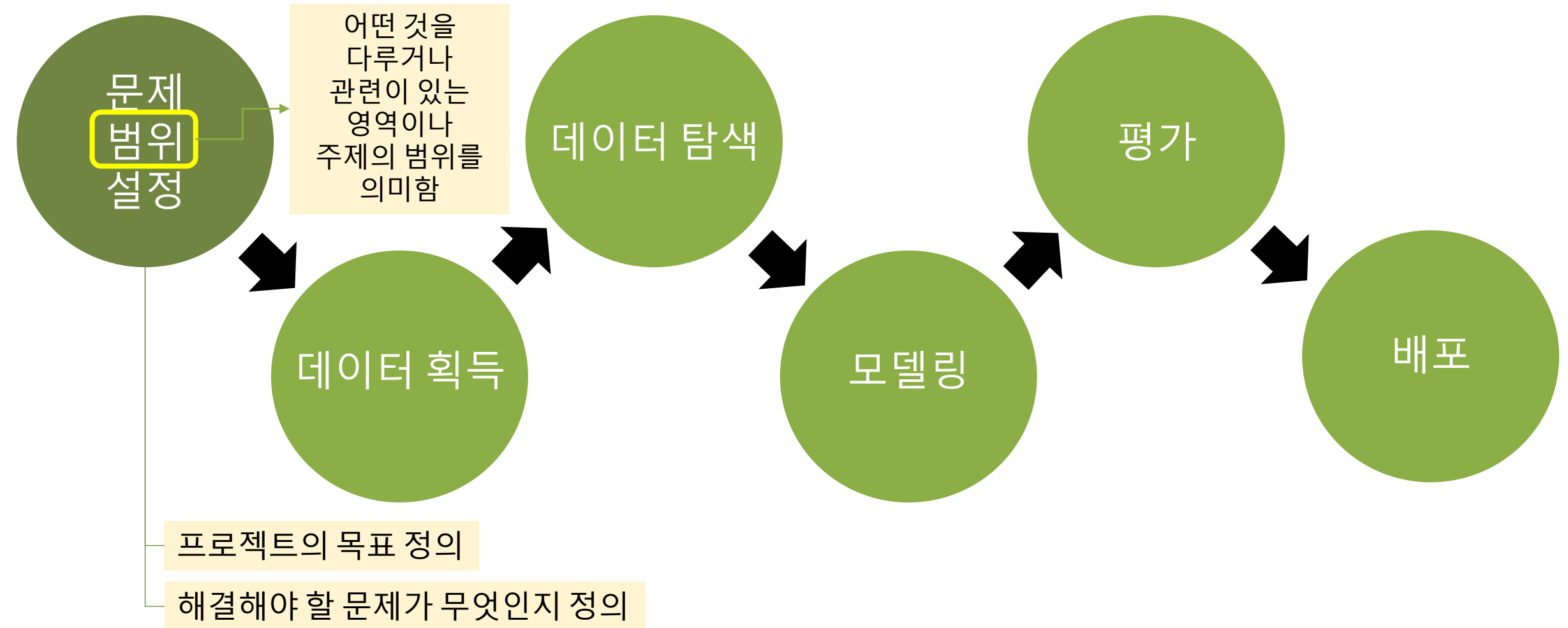


AI 프로젝트 사이클

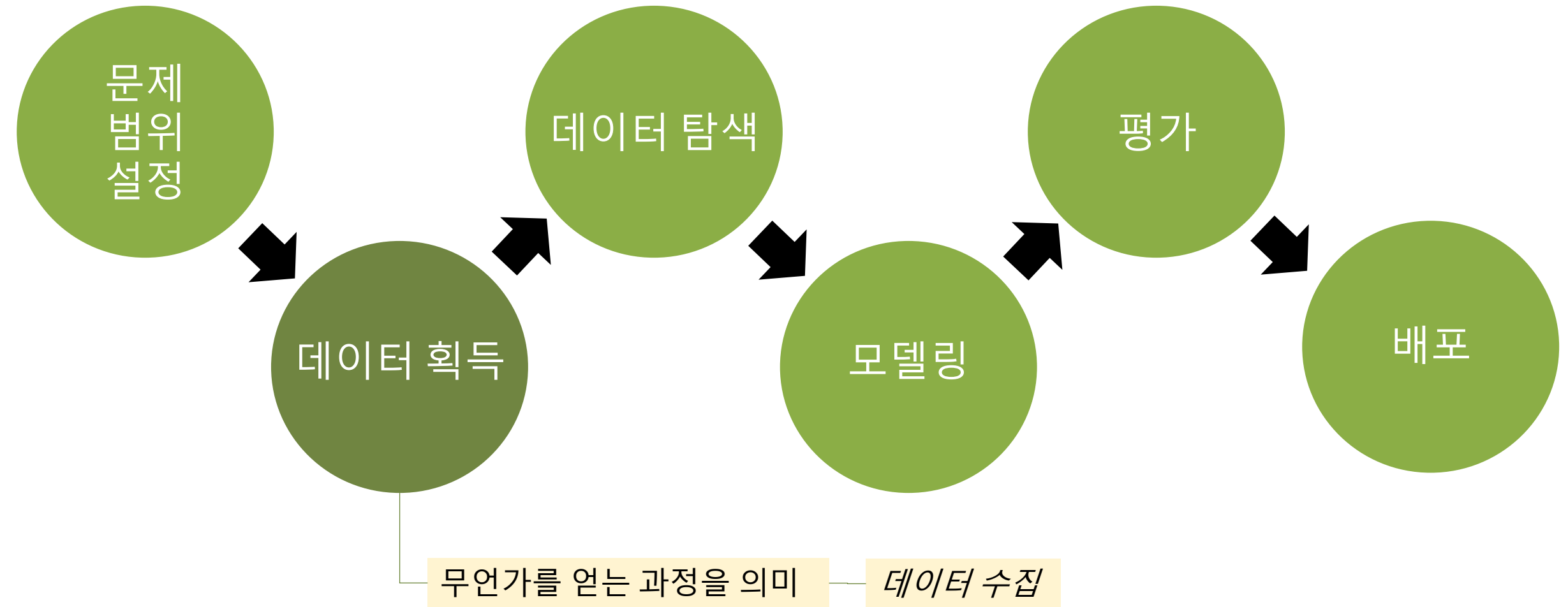
AI 프로젝트 사이클



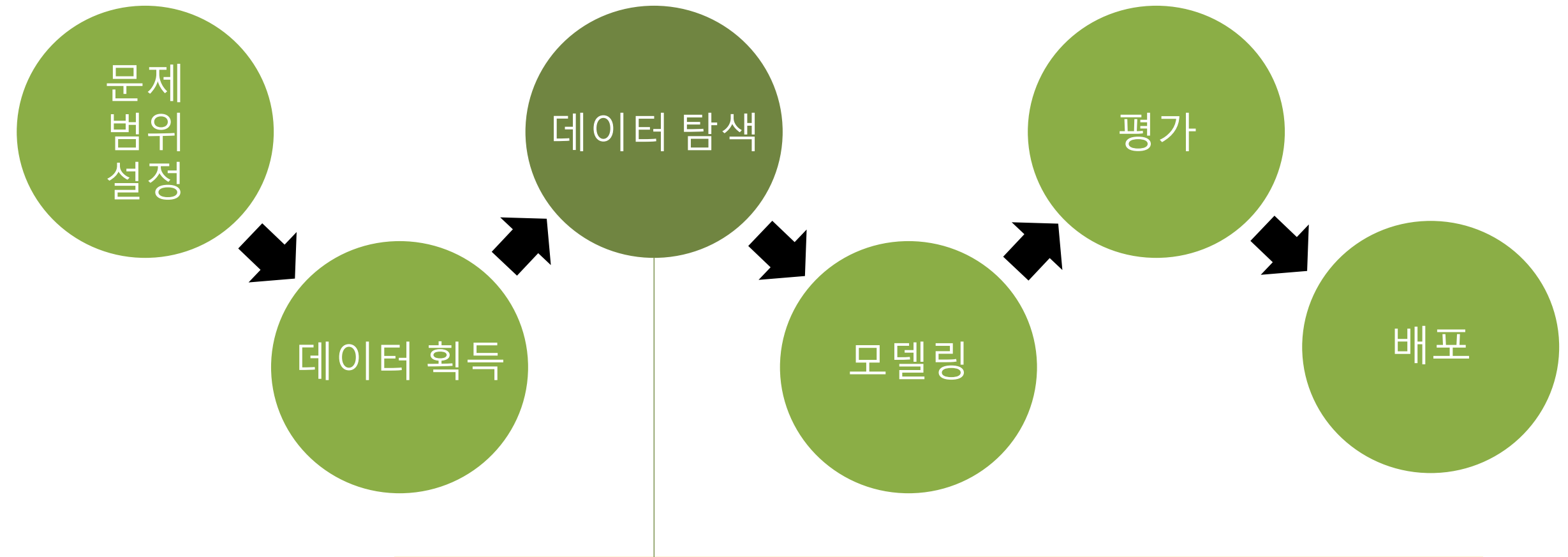
AI 프로젝트 사이클



AI 프로젝트 사이클

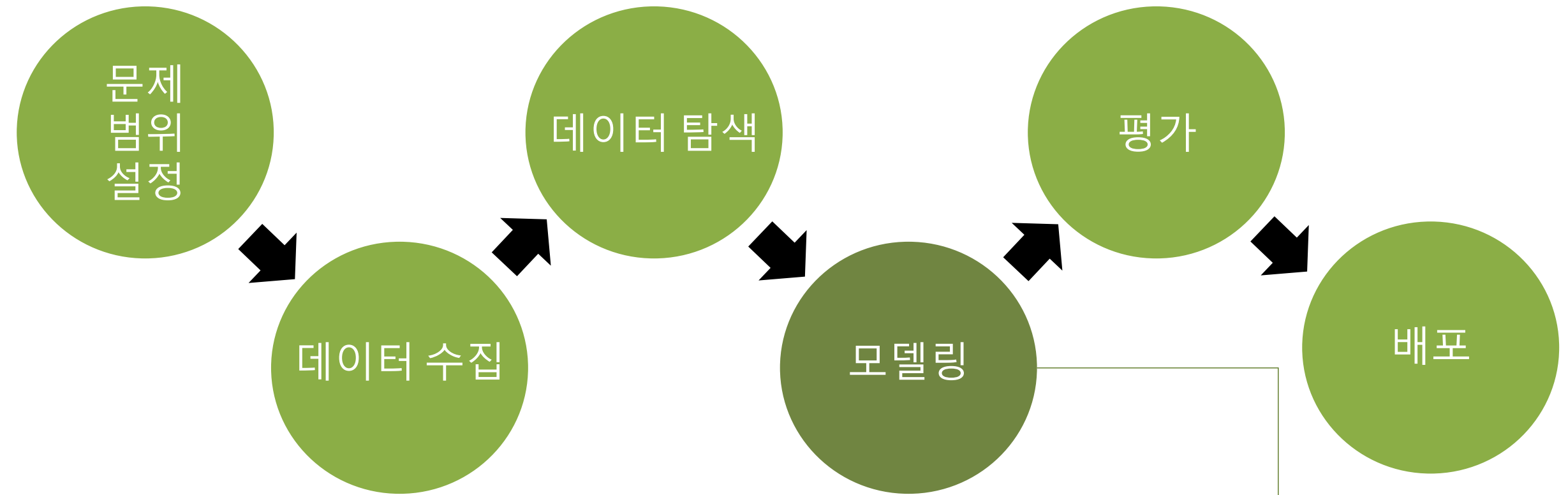


AI 프로젝트 사이클



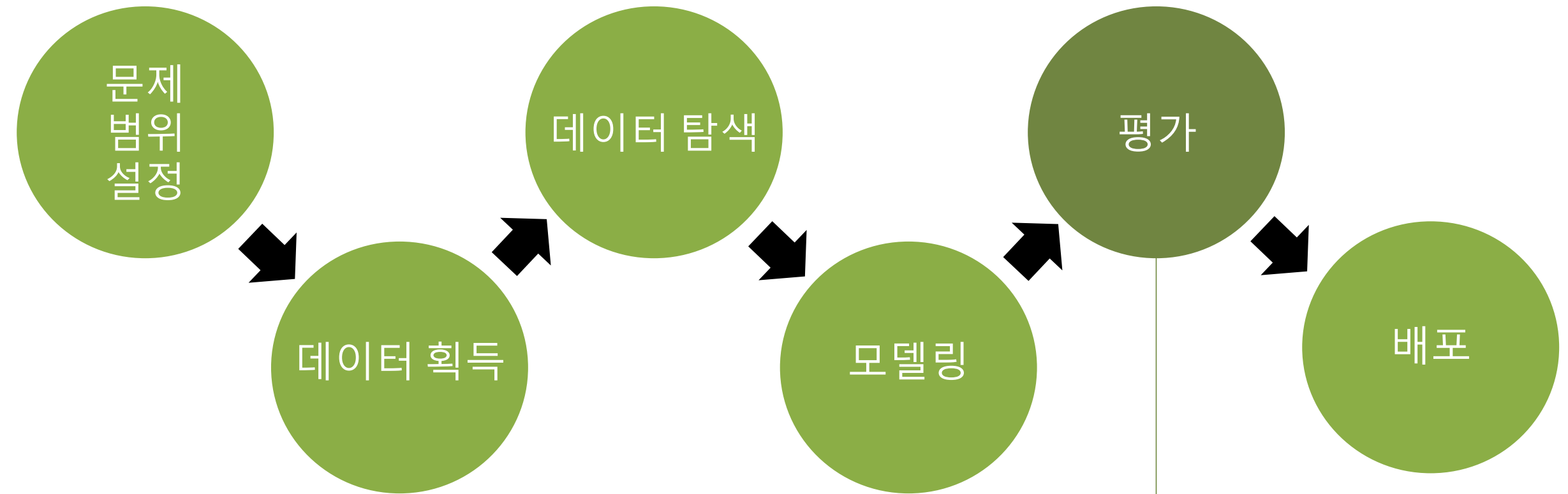
그래프를 통해 또는 사람이 읽을 수 있는 형태로 데이터를 시각화하는 것이 목표

AI 프로젝트 사이클



AI/ML에서 모델은 의사 결정 프로세스를 모방하여 자동화 및 이해 가능
주요 목적은 알고리즘을 구현하고 최상의 결과를 얻을 수 있도록 모델을 미세 조정하는 것

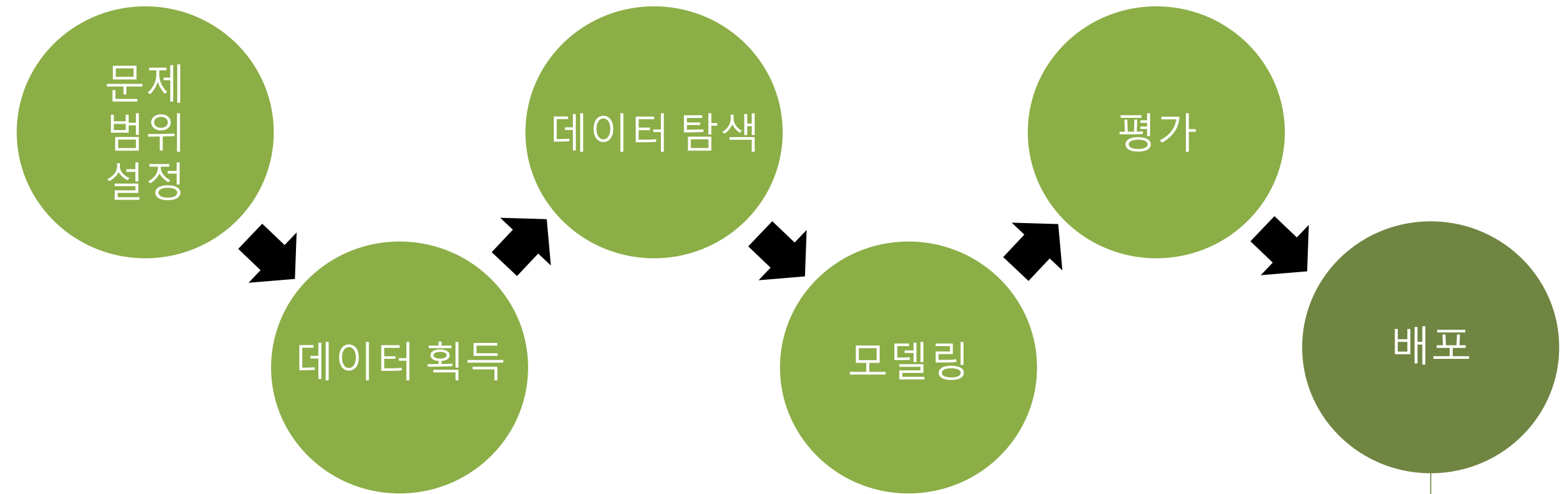
AI 프로젝트 사이클



평가는 어떤 것의 품질, 중요성 또는 가치를 판단하거나 계산하는 것

주요 목적은 다양한 모델을 평가하고 최상의 모델을 선택하는 것

AI 프로젝트 사이클



배포는 유용한 방법으로 무언가를 사용한다는 의미
주요 목표는 선택된 플랫폼에 모델을 기반으로 솔루션을 배포

1. 문제 범위 설정

4W 문제 캔버스

4W 문제 캔버스

누가
WHO ?

무엇을
WHAT ?

어디서
WHERE ?

왜
WHY ?



졸음 모니터 문제 : 버스 운전자들의 졸음 운전

4W 문제 캔버스 – 누가(Who)?

누가 문제를 겪고 있나?

1. 이해 관계인은 누구인가?

- 학생을 포함한 버스 승객
- 버스 운전자

2. 그들에 대해 무엇을 알고 있나?

- 버스 운전자와 학생들은 종종 집에서 학교까지 먼 길을 가야 한다.

4W 문제 캔버스 – 무엇을(What)?

문제의 본질은 무엇인가?

1. 무엇이 문제인가?

- 버스 운전사들의 졸음 운전으로 인해 도로에서 사고가 발생할 수 있습니다.

2. 이것이 문제인지 어떻게 알았는가?

- 버스 운전사의 졸음 운전이 매년 사고로 이어지는 문제임을 보여주는 통계적인 증거가 있습니다.

4W 문제 캔버스 – 어디에서(Where)?

어디에서 문제가 발생하나?

1. 어떤 상황에서 문제가 발생하나? 무엇 때문에 문제가 발생하나?
 - 이 문제는 버스 운전사가 졸리거나 어떤 이유로든 몸이 좋지 않을 때 발생합니다.

4W 문제 캔버스 – 왜(Why)?

왜 여러분은 그것이 해결할 가치가 있는 문제라고 생각합니까?

1. 이해 관계자들에게 중요한 가치는 무엇인가?

- **승객의 안전한 버스 여행을 보장합니다.**

2. 그들의 상황을 어떻게 개선할 수 있을까?

- **그것은 사고 확률을 낮추고 버스 운전자들이 안전하게 운전할 수 있게 해 줄 것입니다.**

문제 범위 지정 양식

우리의	[이해 관계자]	누가
...에 대한 문제가 있다	[이슈, 문제, 니즈]	무엇을
언제/~동안	[맥락, 상황]	어디서
이상적인 해결책은	[해결책의 이점]	왜

문제 범위 지정 양식

졸음 모니터 문제: 버스 운전사들의 졸음 운전

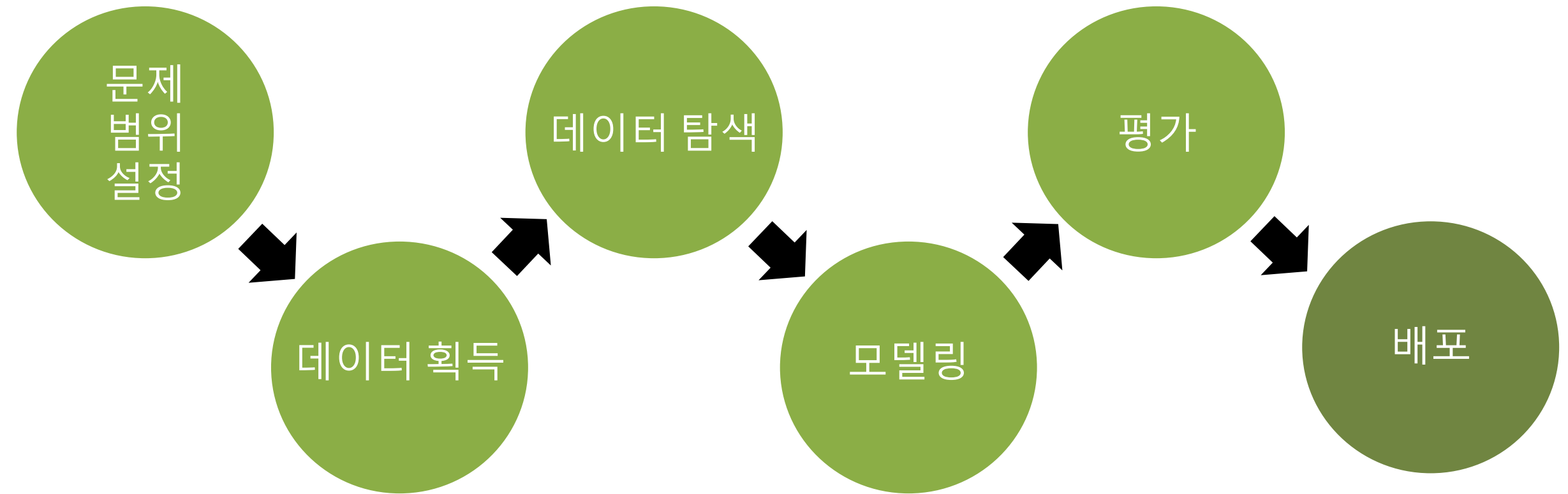
우리의	승객들	누가
...에 대한 문제가 있다	버스 운전사들은 잠이 들 수 있다	무엇을
언제/~동안	긴 버스 여행 중에	어디서
이상적인 해결책은	버스 운전사가 졸릴 때를 감지하는 데 도움이 된다.	왜

문제 범위 지정 양식

전자상거래 플랫폼에서 사기 탐지

우리의	고객들	누가
...에 대한 문제가 있다	그들은 사기를 당했다.	무엇을
언제/~동안	전자 상거래 플랫폼에서 물건을 구매할 때	어디서
이상적인 해결책은	전자 상거래 플랫폼에서 사기를 탐지한다.	왜

AI 프로젝트 사이클



2. 데이터 획득

데이터 소스란 무엇입니까?

- 데이터가 생성되거나 물리적 데이터 소스는 사용 중인 데이터 원본의 위치
- 다른 프로세스가 접근하고 활용하는 한 가장 정제된 데이터
- 데이터베이스, 플랫 파일, 물리적 장치로부터의 실시간 측정, 스크랩된 웹 데이터 또는 인터넷에 널리 퍼져 있는 수많은 정적 및 스트리밍 데이터 서비스

PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
1	0	3	Braund, Mr. Owen Harris	male	22	1	0	A/5 21171	7.25		S
2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Thayer)	female	38	1	0	PC 17599	71.2833	C85	C
3	1	3	Heikkinen, Miss. Laina	female	26	0	0	STON/O2. 3	7.925		S
4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35	1	0	113803	53.1	C123	S
5	0	3	Allen, Mr. William Henry	male	35	0	0	373450	8.05		S
6	0	3	Moran, Mr. James	male			0	330877	8.4583		Q
7	0	1	McCarthy, Mr. Timothy J	male	54	0	0	17463	51.8625	E46	S
8	0	3	Palsson, Master. Gosta Leonard	male	2	3	1	349909	21.075		S
9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27	0	2	347742	11.1333		S
10	1	2	Nasser, Mrs. Nicholas (Adele Achem)	female	14	1	0	237736	30.0708		C
11	1	3	Sandstrom, Miss. Marguerite Rut	female	4	1	1	PP 9549	16.7	G6	S
12	1	1	Bonnell, Miss. Elizabeth	female	58	0	0	113783	26.55	C103	S
13	0	3	Saunderscock, Mr. William Henry	male	20	0	0	A/5. 2151	8.05		S
14	0	3	Andersson, Mr. Anders Johan	male	39	1	5	347082	31.275		S
15	0	3	Vestrom, Miss. Hulda Amanda Adolfina	female	14	0	0	350406	7.8542		S
16	1	2	Hewlett, Mrs. (Mary D Kingcome)	female	55	0	0	248706	16		S
17	0	3	Rice, Master. Eugene	male	2	4	1	382652	29.125		Q
18	1	2	Williams, Mr. Charles Eugene	male			0	244373	13		S
19	0	3	Vander Planke, Mrs. Julius (Emelia Maria Vandemoortele)	female	31	1	0	345763	18		S
20	1	3	Masselmani, Mrs. Fatima	female			0	2649	7.225		C

타이타닉 데이터 세트(출처: Kaggle)


데이터 소스의 예입니다.



The screenshot shows the Kaggle website interface. The left sidebar has a menu with options: Home, Compete, Data (highlighted), Notebooks, Discuss, Courses, and More. The main content area is titled 'Datasets' and includes a search bar, a 'New Dataset' button, and a 'Create Public Datasets' section. The 'Create Public Datasets' section has a button labeled 'Create Public Dataset' and an illustration of two people working on a laptop. Below this, there is a search bar for datasets, a 'Feedback' button, and a 'Filter' button. A list of public datasets is displayed, including 'US Election 2020', 'US Election 2020 Tweets', and 'COVID-19 data from John Hopkins University'. On the right, there is a section for 'Open Tasks' with links to various competitions like 'Predict U.S. Airbnb Prices' and 'Mt Cars'.

데이터 획득 관련 Kaggle 웹사이트

Amazon Web Services (AWS)

Registry of Open Data on AWS

About

This registry exists to help people discover and share datasets that are available via AWS resources. [Learn more about sharing data on AWS.](#)

See [all usage examples for datasets listed in this registry.](#)

See datasets from [Facebook Data for Good](#), [NASA Space Act Agreement](#), [NIH STRIDES](#), [NOAA Big Data Program](#), [Space Telescope Science Institute](#), and [Amazon Sustainability Data Initiative](#).

Search datasets (currently 203 matching datasets)

Search datasets

Add to this registry

If you want to add a dataset or example of how to use a dataset to this registry, please follow the instructions on the [Registry of Open Data on AWS GitHub repository](#).

Unless specifically stated in the applicable dataset documentation, datasets available through the Registry of Open Data on AWS are not provided and maintained by AWS. Datasets are provided and maintained by a variety of third parties under a variety of licenses. Please check dataset licenses and related documentation to determine if a dataset may be used for your application.

The Cancer Genome Atlas

[cancer](#) [genomic](#) [life sciences](#) [STRIDES](#) [whole genome sequencing](#)

The Cancer Genome Atlas (TCGA), a collaboration between the National Cancer Institute (NCI) and National Human Genome Research Institute (NHGRI), aims to generate comprehensive, multi-dimensional maps of the key genomic changes in major types and subtypes of cancer. TCGA has analyzed matched tumor and normal tissues from 11,000 patients, allowing for the comprehensive characterization of 33 cancer types and subtypes, including 10 rare cancers. The dataset contains open Clinical Supplement, Biospecimen Supplement, RNA-Seq Gene Expression Quantification, miRNA-Seq Isoform Expression Quantificati...

[Details →](#)

Usage examples

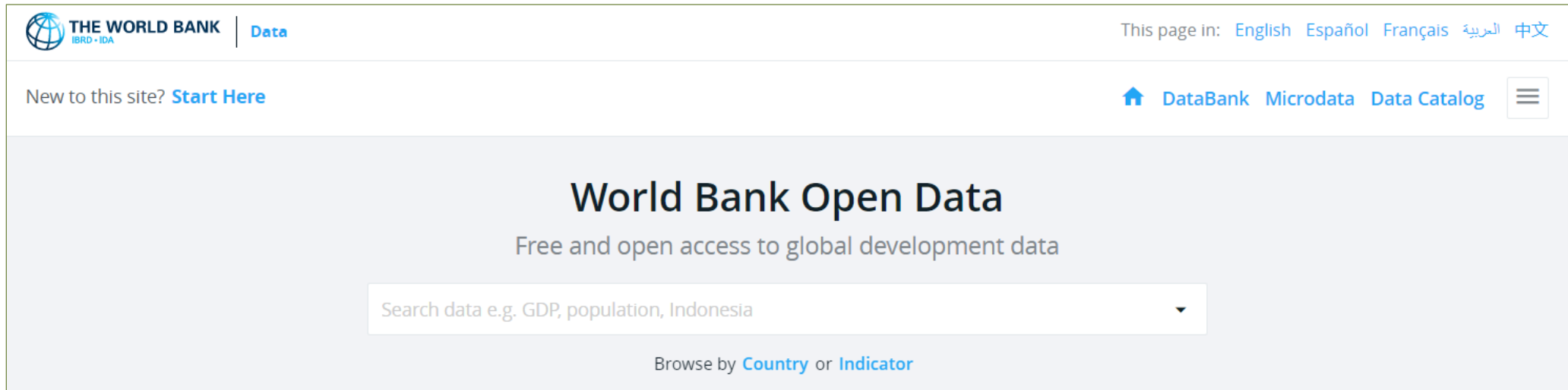
- [Comprehensive Analysis of Alternative Splicing Across Tumors from 8,705 Patients](#) by André Kahles, Kjong-Van Lehmann, et al.
- [Scalable Open Science Approach for Mutation Calling of Tumor Exomes Using Multiple Genomic Pipelines](#) by Kyle Ellrott, Matthew H. Bailey, et al.
- [Pan-Cancer Analysis of lncRNA Regulation Supports Their Targeting of Cancer Genes in Each Tumor Context](#) by Hua-Sheng Chiu, Sonal Somvanshi, et al.
- [Oncogenic Signaling Pathways in The Cancer Genome Atlas](#) by Francisco Sanchez-Vega, Marco Mina, et al.
- [ISB Cancer Genomics Cloud](#) by Institute for Systems Biology

[See 29 usage examples →](#)

데이터 획득 관련 AWS 웹사이트

The World Bank

GNP, 교육, 보건, 영양 등 다양한 국가의 주요 지표를 이용하여 세계 발전에 대한 공개 데이터를 보유



데이터 획득 관련 World Bank 웹사이트

미국(USA) 정부

미국에 관한 자료



The screenshot shows the USA.gov homepage. At the top is the USA.gov logo, a search bar labeled 'Search All Government', and the phone number '1-844-USA-GOV1'. Below the search bar is a navigation menu with links: 'All Topics and Services', 'Benefits, Grants, Loans', 'Government Agencies and Elected Officials', 'Jobs and Unemployment', 'Money and Taxes', 'Travel and Immigration', and 'Voting and Elections'. A 'Español' button is located on the right side of the page. The main content area features a section titled '< About the U.S.' with a list of links: 'American Flag', 'Branches of the U.S. Government', 'Budget of the U.S. Government', 'Data and Statistics about the U.S.' (highlighted with a red bar), and 'Government Agencies and Elected Officials'. To the right of this list is a large section titled 'Data and Statistics about the U.S.' in red text. Below this title is a paragraph: 'Find data about the U.S., such as demographic and economic data, population, and maps. Get information about the 2020 U.S. Census and learn how to respond.' Below the paragraph is a light blue box titled 'On This Page' containing four links: 'Respond to the 2020 Census', 'Find Data and Statistics from the Government', 'U.S. Census Data and Statistics', and 'Maps'.

usa.gov

Search All Government

1-844-USA-GOV1

All Topics and Services

Benefits, Grants, Loans

Government Agencies and Elected Officials

Jobs and Unemployment

Money and Taxes

Travel and Immigration

Voting and Elections

Español

< About the U.S.

American Flag

Branches of the U.S. Government

Budget of the U.S. Government

Data and Statistics about the U.S.

Government Agencies and Elected Officials

Data and Statistics about the U.S.

Find data about the U.S., such as demographic and economic data, population, and maps. Get information about the 2020 U.S. Census and learn how to respond.

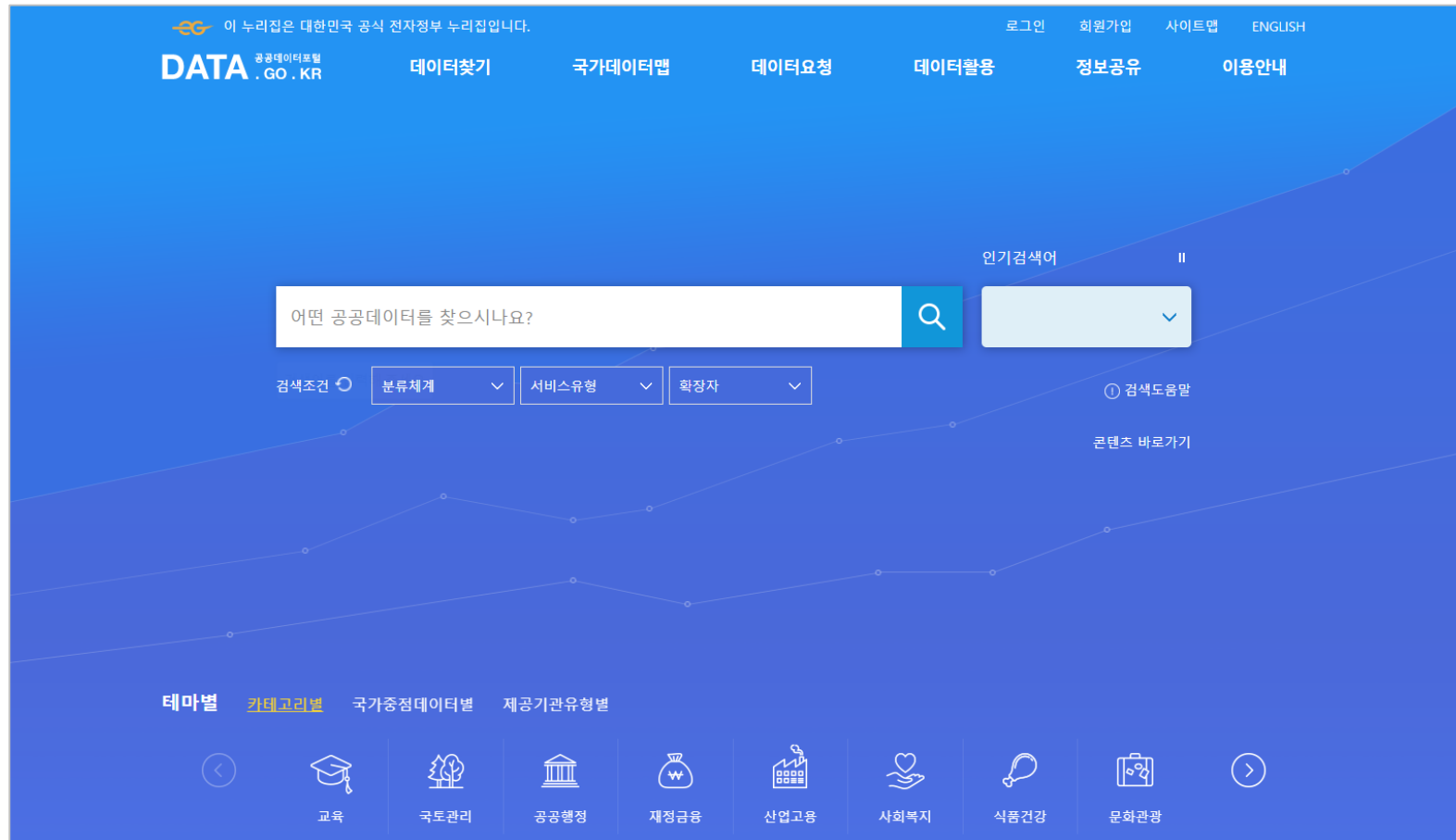
On This Page

- [Respond to the 2020 Census](#)
- [Find Data and Statistics from the Government](#)
- [U.S. Census Data and Statistics](#)
- [Maps](#)

데이터 획득 관련 U.S. Gov 웹사이트

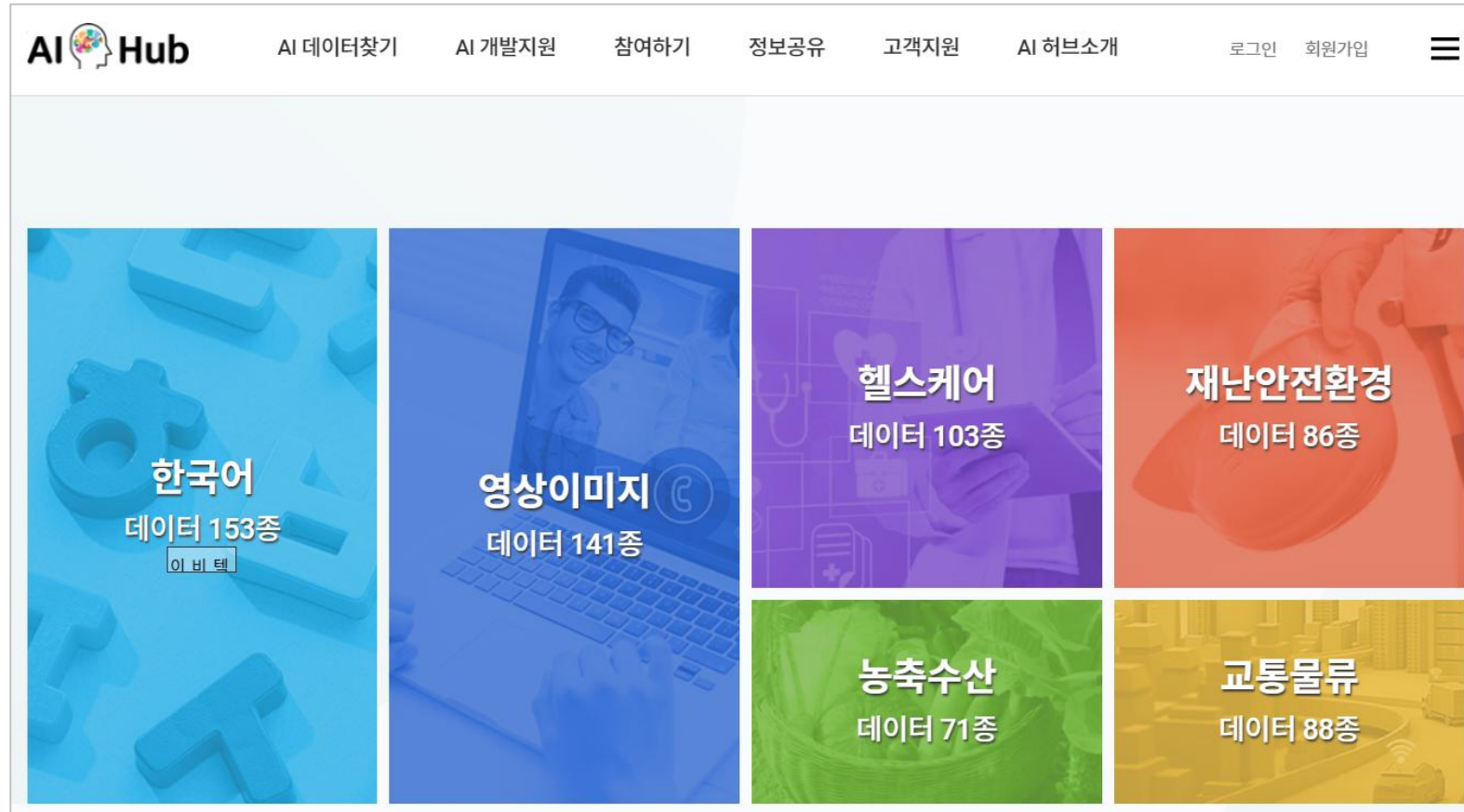
공공 데이터 포털

대한민국 공식 전자정부 누리집



[데이터 획득 관련 대한민국 정부 웹사이트](#)

공공 데이터 포털



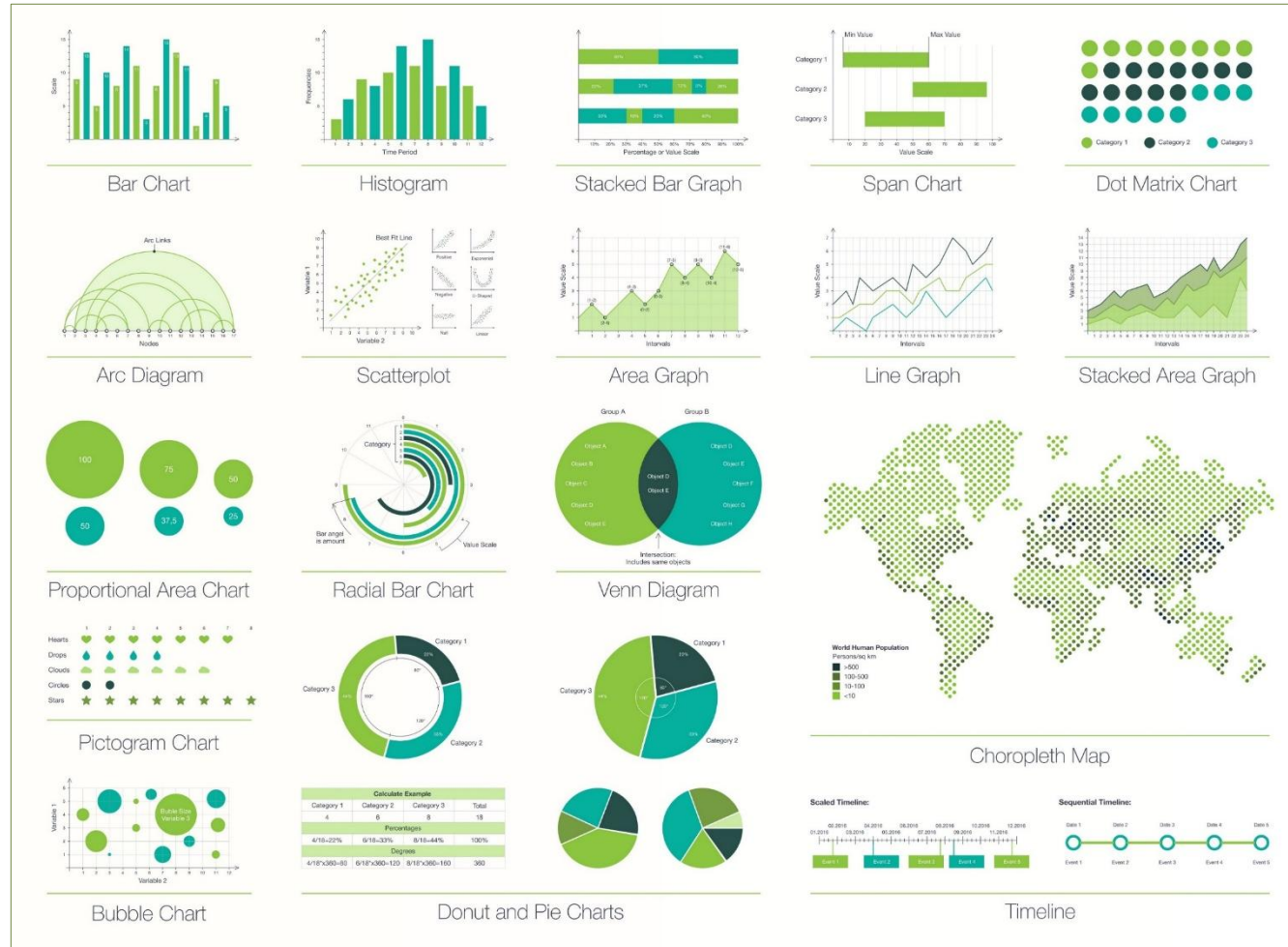
지능정보산업 인프라 조성사업으로 추진한 AI 학습용 데이터(6개 분야)와 국내외 기관/기업에서 보유한 AI 학습용 데이터를 공개

3. 데이터 탐색

데이터 시각화 도구



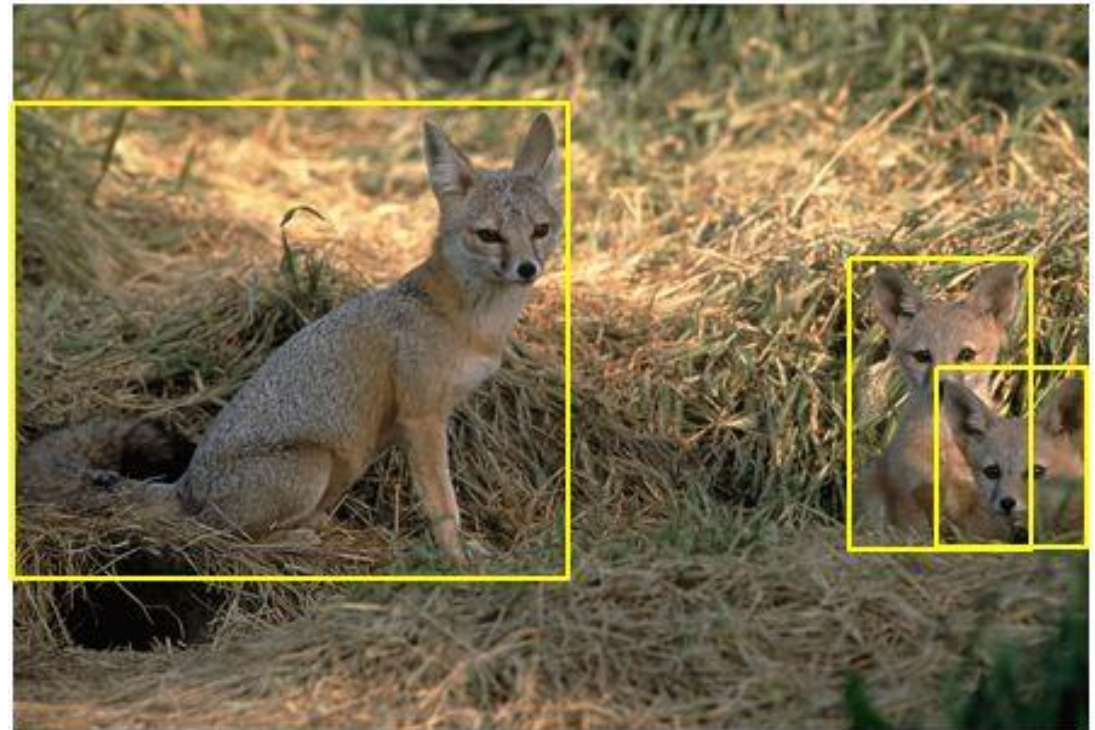
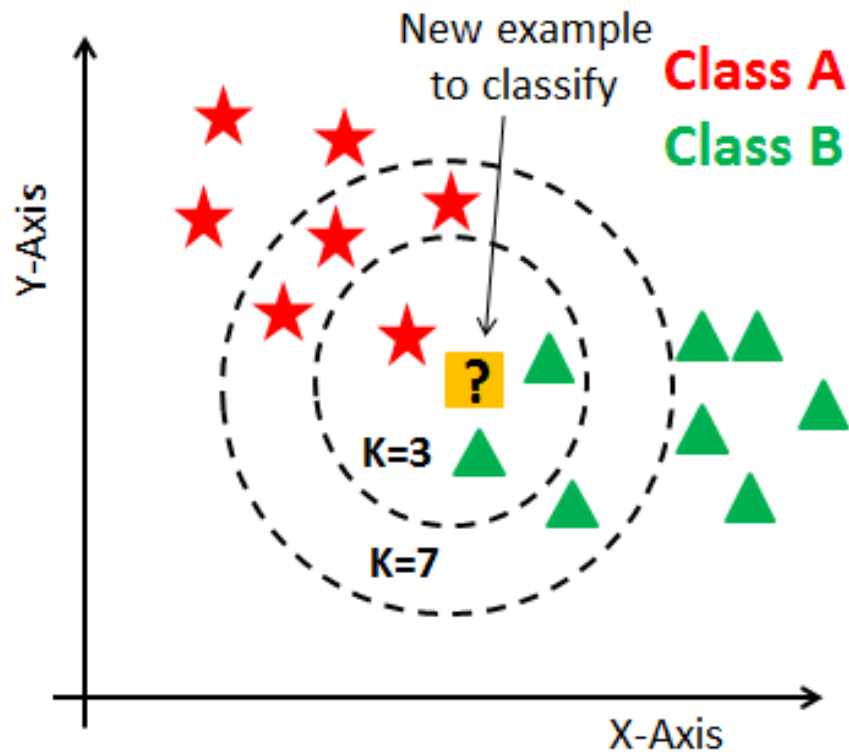
데이터의 시각적 표현



4. 모델링

모델링 기법

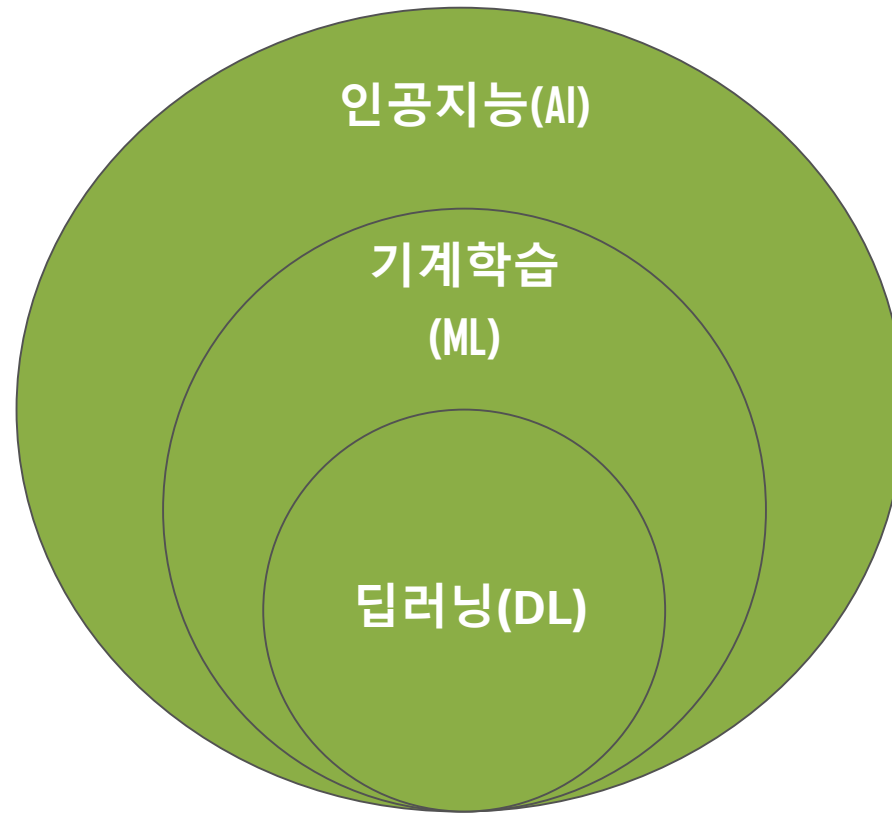
모델링 : 데이터에서 패턴을 찾는 데 도움이 되는 과정



좋은 모델을 만드는 과정은 훈련을 통해 이루어짐

훈련 : 데이터에 가장 적합하게 모델을 미세 조정하는 방법

AI, ML 및 DL 의 관계

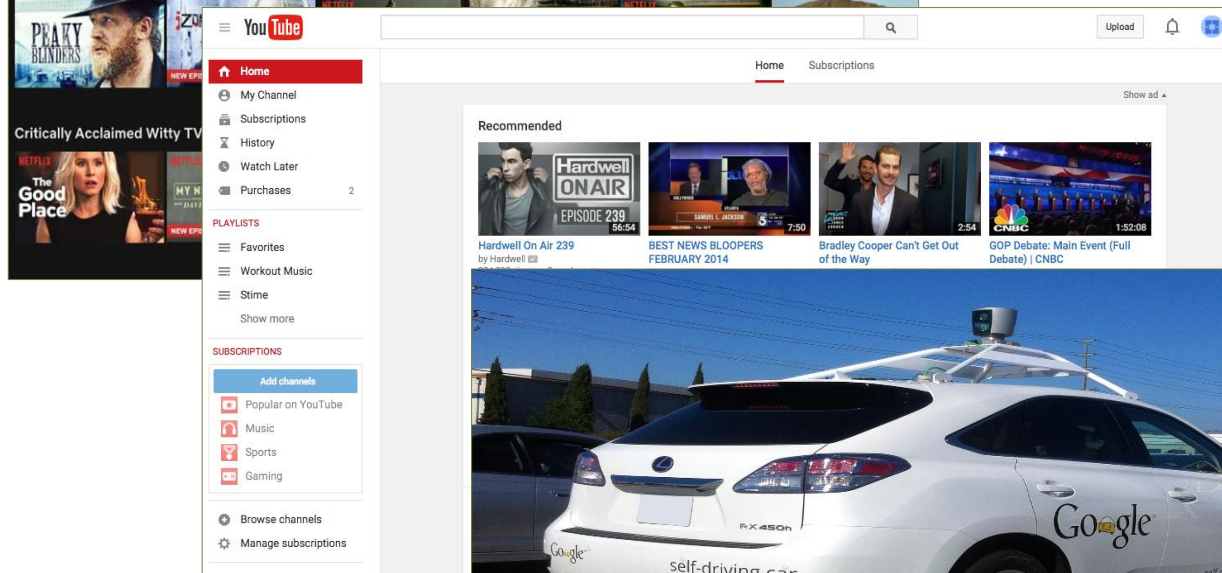
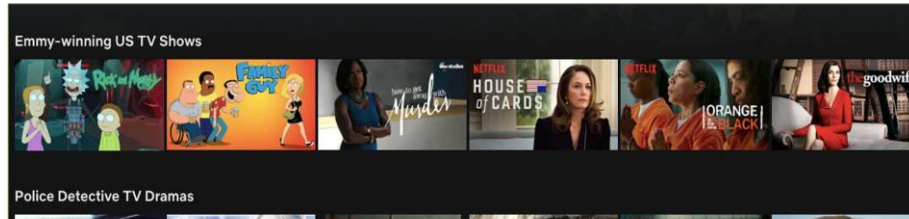


AI, ML 및 DL의 예

1. AI - 컴퓨터가 인간의 지능을 흉내 낼 수 있게 하는 기술

2. 기계학습 - AI의 하위 집합으로 기계가 경험이 있는 업무를 개선할 수 있도록 함

3. 딥 러닝- 기계학습과 AI의 하위 집합으로 소프트웨어가 방대한 양의 데이터로 작업을 수행하도록 스스로 훈련할 수 있음



어떤 기계학습 모델이 있습니까?

일반적으로, 세 가지 계열의 모델이 있습니다.

- 지도 학습
- 비지도 학습
- 강화 학습

ML 모델 계열

```
graph TD; A[ML 모델 계열] --> B[지도 학습]; A --> C[비지도 학습]; A --> D[강화 학습]; B --> E[훈련을 통해 관계 확인]; C --> F[새로운 패턴 찾기]; D --> G[보상 메커니즘을 통한 학습];
```

지도 학습

훈련을 통해 관계
확인

비지도 학습

새로운 패턴 찾기

강화 학습

보상 메커니즘을
통한 학습

지도 학습이란 무엇인가?

- 지도 학습은 알려진 데이터와 레이블이 지정된 데이터에서 학습하는 알고리즘입니다.
- 알 수 없는 데이터가 모델에 입력되면 학습한 내용을 바탕으로 응답을 얻을 수 있습니다.

실제 사례는 다음과 같습니다:

- 생체 인식을 이용한 출석 확인
- 사기 탐지

비지도 학습이란 무엇인가?

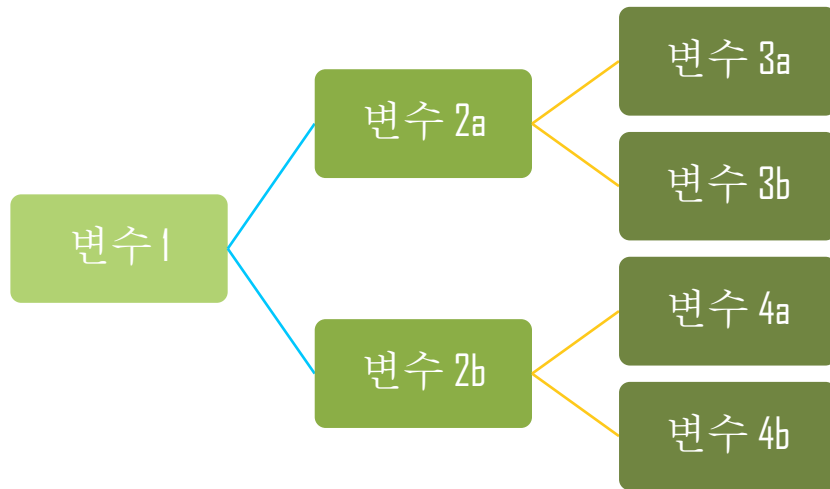
- 비지도 학습은 감독 없이 알려지지 않은 데이터로부터 학습하는 것입니다.
- 기계는 라벨이 부착되지 않은 데이터에서 패턴을 식별하고 적절한 결과를 실행할 수 있습니다.

실제 사례는 다음과 같습니다:

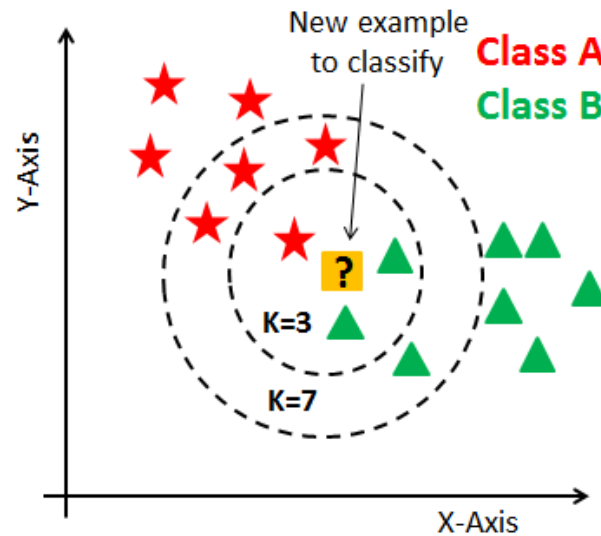
- 사고가 발생하기 쉬운 영역 식별
- 인터넷 쇼핑몰 장바구니 분석

일반적인 지도 학습 알고리즘 예시

분류
(범주형 데이터)

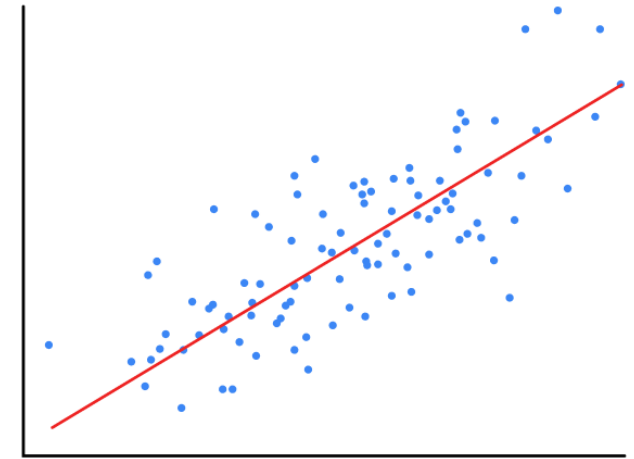


의사 결정 트리



K-NN

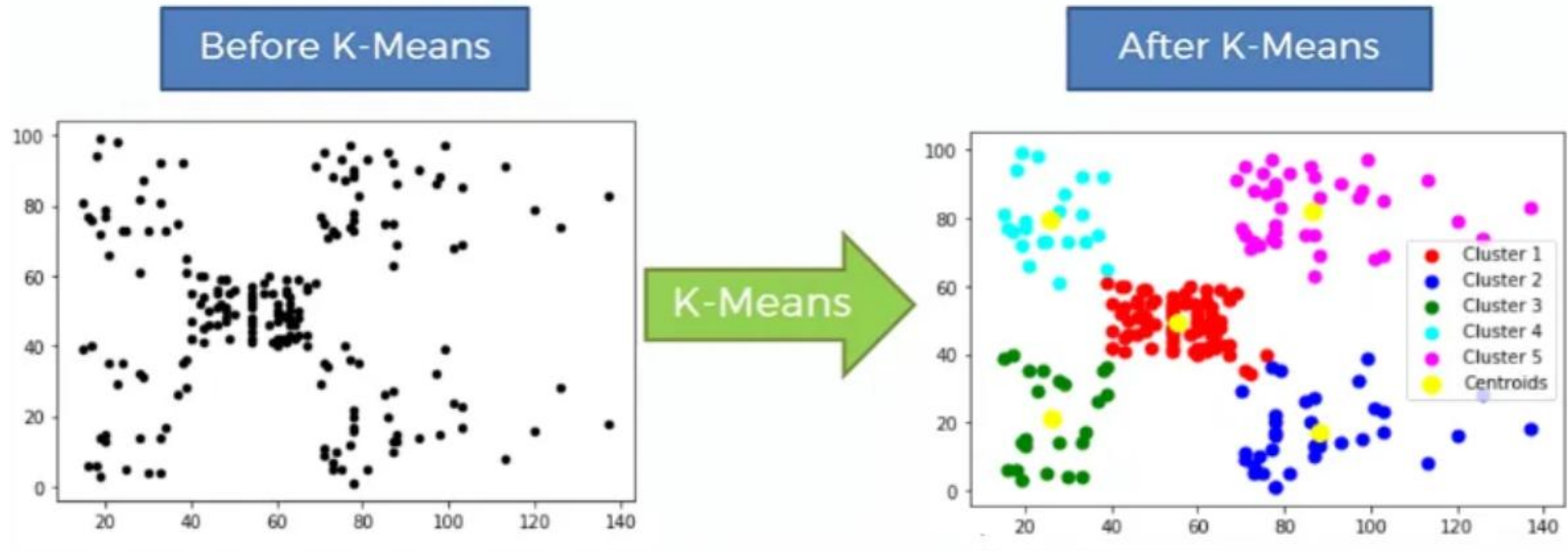
회귀
(숫자 데이터)



선형 회귀 분석

일반적인 비지도 학습 알고리즘 예제

군집화



<https://ai.plainenglish.io/what-is-k-means-clustering-3060791cb589>

K-평균 군집화

5. 평가

모델에 대한 평가 측정 지표

분류

- 분류 정확도
- 혼동 행렬

회귀

- 루트 평균 제곱 오차(RSE)
- 평균 절대 오차(MAE)

클러스터링

- 실루엣 방법

6. 배포

배포할 플랫폼



모바일 애플리케이션



웹사이트
애플리케이션



클라우드 서버