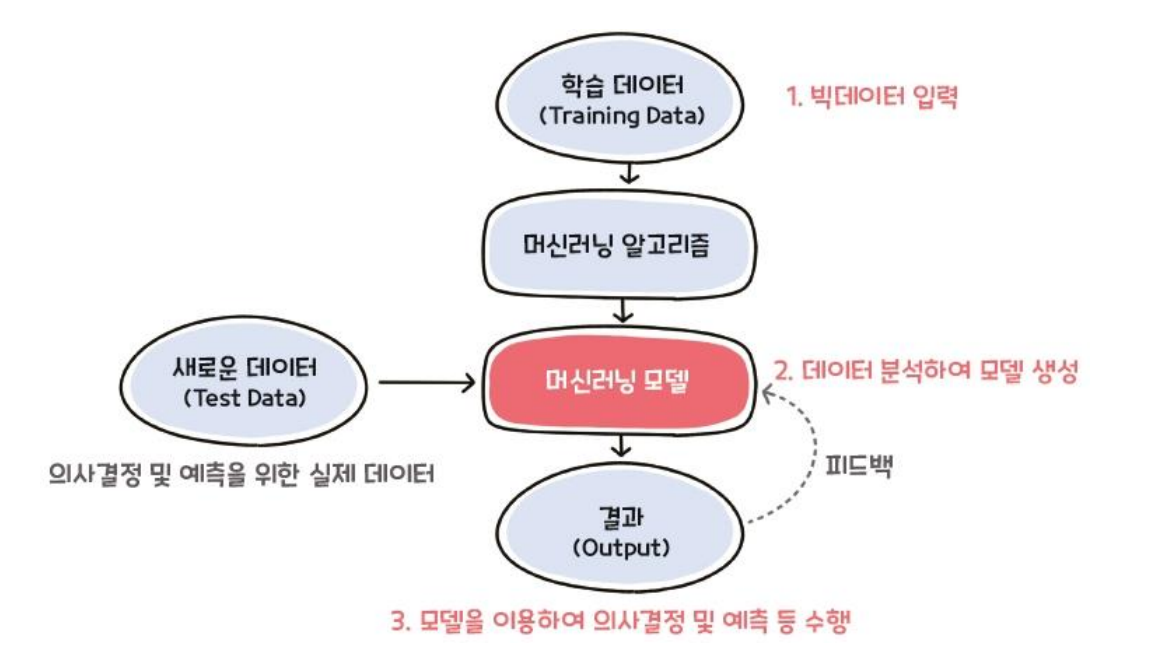
**머신러닝이 뭔가요?**

* 컴퓨터를 인간처럼 학습하게 함으로써 인간의 도움 없이도 컴퓨터 스스로가 새로운 규칙을 발견할 수 있도록 하는 기술
* 기본적으로 알고리즘을 이용해 데이터를 분석, 분석을 통해 학습, 학습한 내용을 기반으로 판단이나 예측을 함

**머신러닝이 스스로 학습하여 데이터를 처리하는 과정**

1. 빅데이터를 입력
2. 데이터를 분석하여 모델 제작
3. 모델을 이용하여 의사결정 및 예측 등을 수행

**머신러닝 학습절차**



**특징추출(Feature Extraction)** : 머신러닝에서 컴퓨터가 스스로 학습하려면 사람이 인지하는 데이터를 컴퓨터가 인지할 수 있는 데이터로 변환해야 하는데, 이 작업을 위해 데이터별로 어떤 특징을 가지는지 찾아 내고 그것을 토대로 데이터를 벡터로 변환하는 것

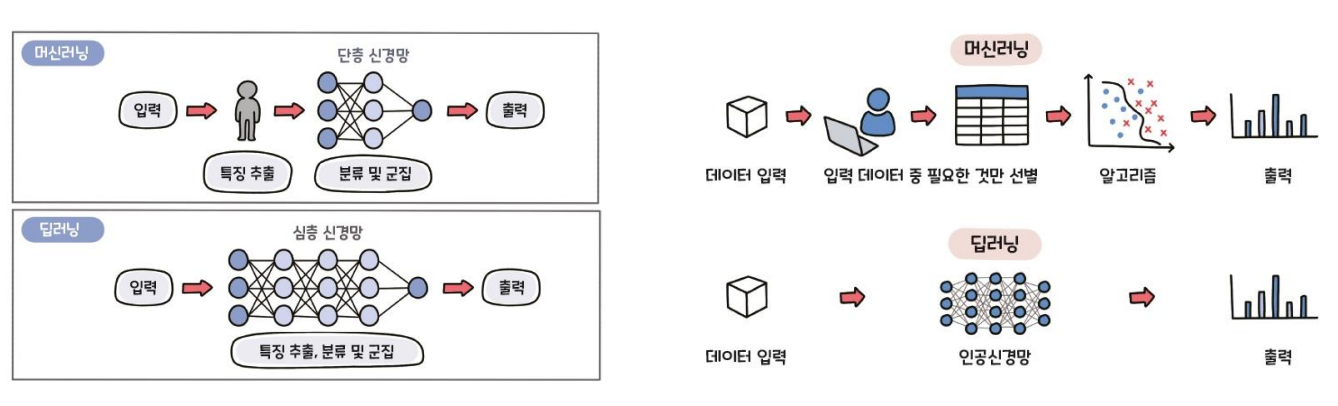
도표이(가) 표시된 사진

자동 생성된 설명

**머신러닝과 딥러닝의 차이점**

**인간의 개입 유무**

* 머신러닝은 사람이 학습 데이터에 레이블(정답)을 알려주거나 데이터의 특징을 추출하는 등 어느 정도 개입
* 딥러닝은 인간의 개입 없이 컴퓨터 스스로 학습함



**데이터 의존도**

* 딥러닝은 주어진 문제를 해결하기 위해 중요한 특징을 직접 추출한다. 그래서 데이터의 양이 충분하지 않으면 정확한 특징을 추출할 수 없다.
* 충분한 양의 데이터가 주어진다면 사람이 인지하지 못한 중요한 특징들까지 찾아낼 수 있을 정도로 좋은 성능을 발휘한다.

도표이(가) 표시된 사진

자동 생성된 설명

**머신러닝 분류**

**지도학습(Supervised Learning)**

학습을 위한 다양한 피처와 분류 결정값인 레이블(Label) 데이터로 모델을 학습한 뒤, 별도의 테스트 데이터 세트에서 미지의 레이블을 예측, 즉 지도학습은 명확한 정답이 주어진 데이터를 먼저 학습한 뒤 미지의 정답을 예측하는 방식

● 학습 데이터 셋 : 학습을 위해 주어진 데이터 셋

● 테스트 데이터 셋 : 머신러닝 모델의 예측 성능을 평가하기 위해 별도로 주어진 데이터 셋

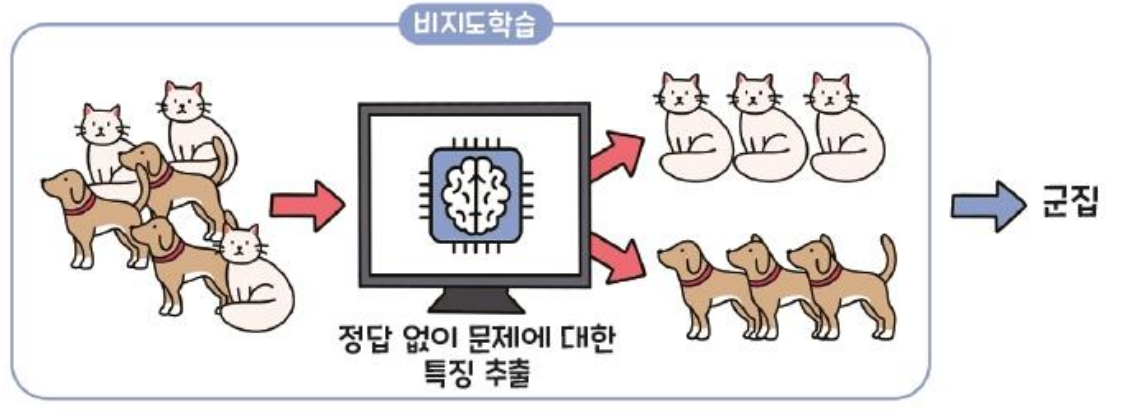
* 문제와 답을 함께 학습함으로써 미지의 문제에 대한 올바른 답을 예측하는 학습
* 지도학습에서 사용하는 모델로는 크게 예측과 분류가 있음

도표이(가) 표시된 사진

자동 생성된 설명

**비지도학습(Unsupervised Learning)**

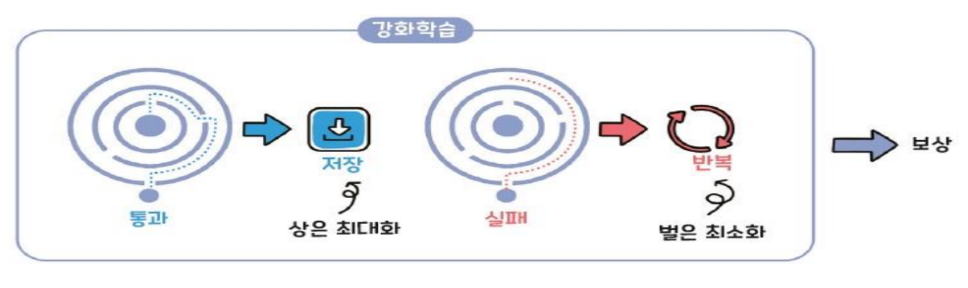
* 지도학습과 다르게 조력자의 도움 없이 컴퓨터 스스로 학습하는 형태
* 컴퓨터가 훈련 데이터를 이용하여 데이터들 간의 규칙성을 찾음



X (입력 데이터)와 y(지도학습에서 레이블)의 관계를 파악했던 지도학습과는 달리, 비지도학습은 x간의 관계를 스스로 파악한다. 지도학습과 다른점은 y(레이블)의 차이이며 비지도학습에서 사용하는 모델로는 군집(Clustering)이 있다

**강화학습(Reinforcement Learning)**

* 자신이 한 행동에 대해 보상을 받으면 학습하는 것
* 컴퓨터가 주어진 상태에 대해 최적의 행동을 선택하도록 학습하는 방법



**강화학습 개념**

Agent(에이전트) : 주어진 문제 상황에서 행동하는 주체

State(상태) : 현재 시점에서의 상황

Action(행동) : 플레이어가 취할 수 있는 선택지

Reward(보상) : 플레이어가 어떤 행동을 했을 때 다라오는 이득

Environment(환경) : 문제 그 자체를 의미

Observation(관찰) : 에이전트가 수집한(보고/듣는) 환경에 대한 정보

* 주어진 환경에서 에이전트가 선택한 행동에 따라 그 행동이 옳은 선택이면 상을 받고, 잘못된 선택이면 벌을 받음
* 강화학습은 에이전트가 상태를 계속 주시하면서 보상이 높은 쪽으로 학습하게 됨

**분류 (Classification)**

* 레이블이 포함된 데이터를 학습하고 유사한 성질을 갖는 데이터끼리 분류한 후, 새로운 입력 데이터가 어느 그룹에 속하는지를 찾아내는 기법

도표이(가) 표시된 사진

자동 생성된 설명

**이진 분류(Binary Classification)** : 데이터를 2개의 그룹으로 분류

**다중 분류(Multiclass Classification)** : 데이터를 3개의 그룹 이상으로 분류

도표이(가) 표시된 사진

자동 생성된 설명