

## 머신러닝의 개념

### [머신러닝의 개념]

- 머신러닝
  - 애플리케이션 수정 없이도 데이터를 기반으로 패턴을 학습하고 추론하는 알고리즘 기법
  - 데이터 마이닝, 영상인식, 음성인식, 자연어 처리에서 머신러닝 적용 -> 급속하게 발전

### [머신러닝 왜 필요한가?]

- 머신러닝은 복잡한 문제를 데이터 기반으로 숨겨진 패턴을 인지해 해결
- 머신러닝 알고리즘은 통계적 신뢰도 강화, 예측 오류 최소화
- 다양한 수학적 기법 적용 -> 데이터 내의 패턴 스스로 인지, 신뢰도 있는 결과 도출

### [머신러닝의 분류]

- 지도학습
  - 명확한 결정값이 주어진 데이터 학습
  - Ex) 분류, 회귀, 시각/음성감지/인지
- 비지도학습
  - 결정값이 주어지지 않은 데이터 학습
  - Ex) 군집화(클러스터링), 차원 축소

### [머신러닝 기반의 예측 분석]

데이터 분석 영역은 머신러닝 기반의 예측 분석으로 재편

### [머신러닝 알고리즘 유형]

- 기호주의: 결정 트리 등
- 연결주의: 신경망/딥러닝
- 유전 알고리즘
- 베이지안 통계
- 유추주의: KNN, 세포트 벡터 머신

### [머신러닝의 단점]

- 데이터에 너무 의존적 (Garbage In, Garbage Out) -> 편향된 데이터가 편향된 결과 도출
- 학습 시 최적의 결과 도출을 위해 수립된 머신러닝 모델은 실제 환경에서 과적합되기 쉬움
- 복잡한 머신러닝 알고리즘으로 도출된 결과에 대해 논리적인 이해가 어려울 수 있음
- 데이터만 넣으면 자동으로 최적화된 결과를 도출할 것이라는 것은 환상

### [왜 데이터 수집에 열광하는가?]

다양하고 광대한 데이터를 기반으로 만들어진 머신러닝 모델은 더 좋은 품질을 약속