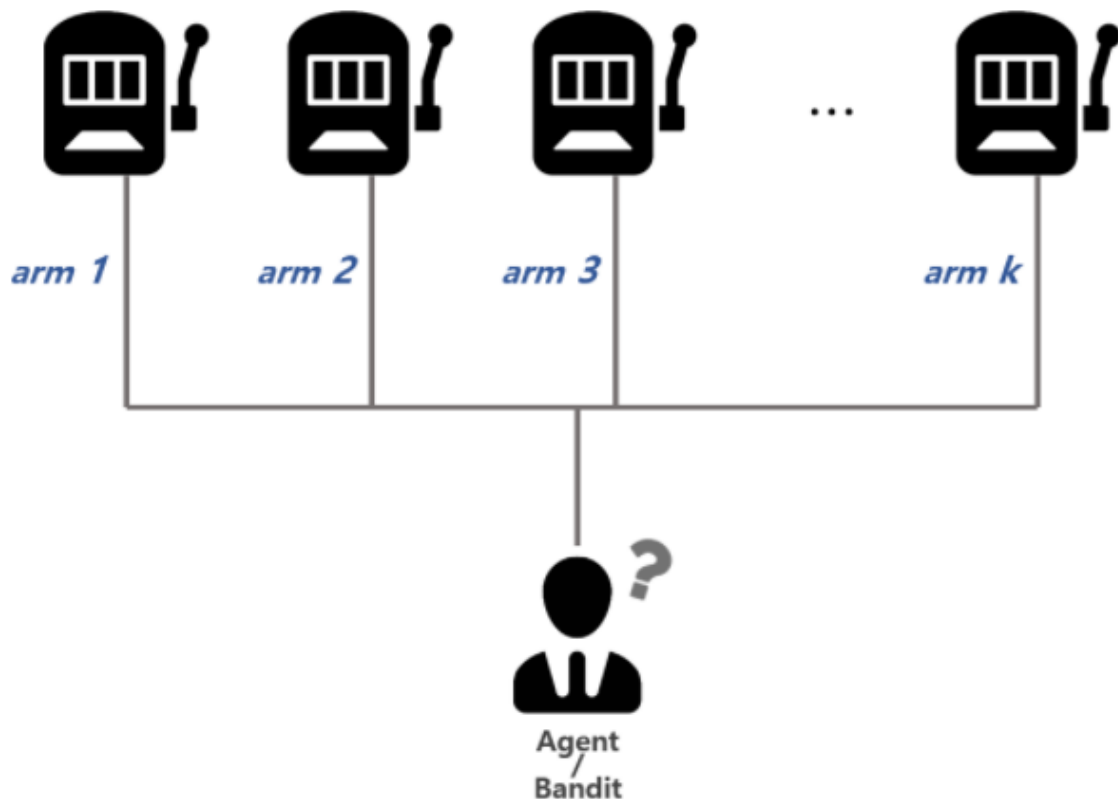


# Multi Armed Bandit

직역 : 팔이 많은 강도?



**일정 시간(H)** 동안 슬롯 머신을 무제한으로 돌릴 수 있는 기회가 팔이 여러 개인 강도에게 주어졌다고 가정하자. 여기서 강도는 **한번에 하나의 팔만을 이용해** 슬롯 머신을 이용할 수 있다. 또한 당연히 강도는 여기서 가장 많은 이득을 주는 방식을 선택해야 한다.

하지만 여기서 의문점이 하나 생기는데

"제일 처음 시작할 땐 아무런 정보가 없는데 어떤 기준으로 팔을 선택해야 하는가?"이다.

물론 처음에는 아무런 정보가 없기 때문에 어떤 선택지를 선택할 지 전혀 알 수가 없다.  
여기서 탐험(Exploration) & 이용(Exploitation)의 개념이 등장한다.

Exploration : 정보 수집을 위해 무작위로 선택지를 고르는 것

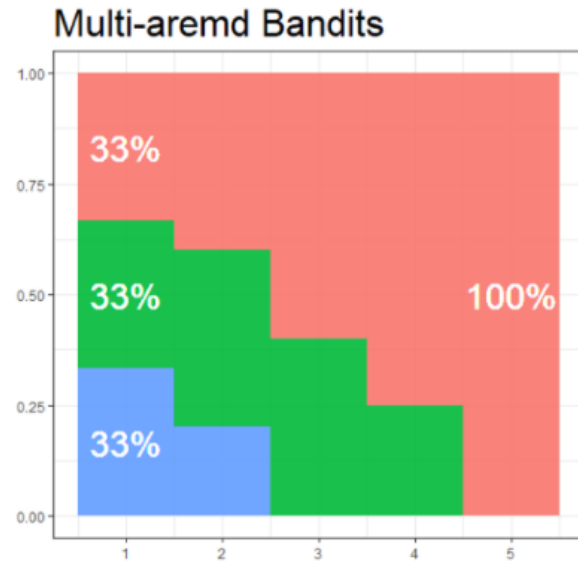
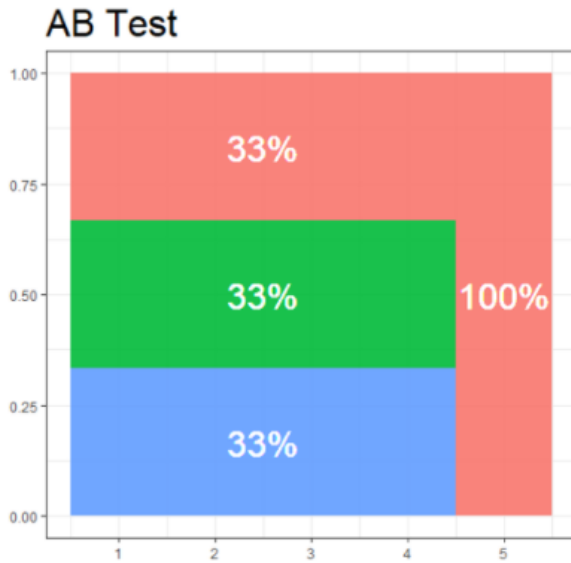
Exploitation : 충분한 정보를 수집한 후 최적의 선택지를 따르는 것

**\*Exploration을 너무 과도하게 하면 무작위로 가는 round 수가 많아져 보상의 양이 적어지게 될 것이며 반대로 Exploitation을 너무 과도하게 하면 우연히 좋았던 선택지만을 따르기 때문에 총 보상에 손실이 발생하게 된다.**



이러한 이유 때문에 각종 알고리즘을 이용해 성능이 좋은 알고리즘을 선택해야 한다.

마케팅 분야의 경우 진행하는 A/B TEST 대신 MAB를 사용하기 위한 목적으로 사용된다.(그림 참조)



## MAB에 등장하는 용어

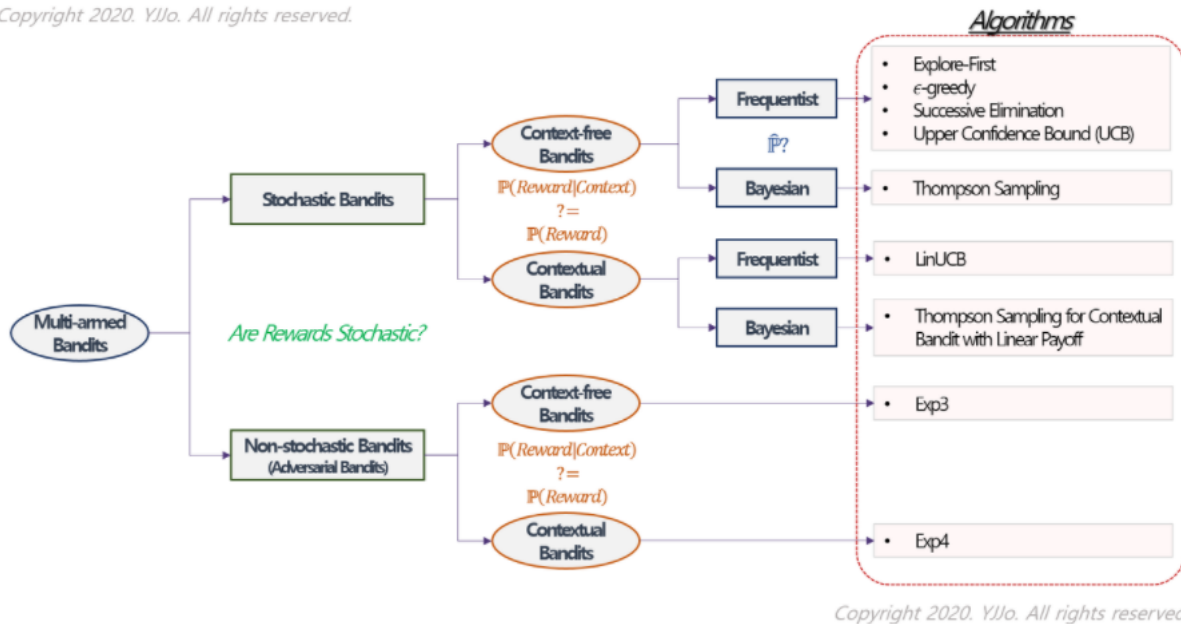
행동의 주체 = (Agent, Bandit, Player) \*모두 같은 뜻임

Episode = (Time-step, Round)

행동 = (Action, Arm)

사전적 정보 = Context

**\*Reward를 어떻게 다루고 어떻게 해석하는지에 따라서 여러가지 Bandit 알고리즘이 존재한다.**



Stochastic & Non-Stochastic 의 차이 - Context(사전 정보)의 유무

Frequentist & Bayesian 관점의 차이

**\*Non-Stochastic은 확률분포를 가정하지 않기 때문에 해당사항이 없음**

## Contextual Bandits

매 Round의 최상의 의사결정 전 관측 가능한 정보를 활용하여 최적의 의사결정을 하는 것.

ex) 비교적 싼 화장품 마케팅 광고를 한다고 할 때 Target을 여성, 20~30대로 잡는 것이 더 효율적이라는 당연한 생각이 드는 것과 같은 원리 (Context 정보는 여기서 **성별, 나이대**)