K-arm Bandit:

*) K-actions, with rewards for each action.
Goal is to find the best k-actions.

Problem I Chose: Recommedation System

*) A System for choosing cricket players based on other situations & states

Here Reward is computed based on result of the selected players

*) We balance exploration & exploitation

The bandit learns which players perform well under different scenarios (like Target, RR) using R. Learning.

*) Each arm represents a cricket player

Factors:

*) Pitch conditions
*) Weather
*) Match Situation (Target, PP)

Reward:

*) How well, the selected player performs

Algorithms to try: → HAVE TO EXPLORE abt this

*) Epsilon-Greedy
*) Upper Conf. Bound
*) Contextual Bandit
*) Thompson Sampling