**Reinfocement Learaning Methods:**

1. MDP(Markov Decision process)
2. MABP(Multi Armed bandit problem)

**Markov Decision Process:**

* MDP is a method used to calculate the values and polices in the state the agent will travel to reach the goal.
* MDP has multiple states, MDP are basically used for applications like building a AI game or in Robotics where the agent should learn what is the value for each step it takes.
* MDP can also be used for recommednation system but most of the recommendation system does not require that high level of computation, instead Recommendation system can be built using algorithms for MABP.

**Multi armed badit:**

* MABP is common RL problem where a gambler uses multiple slot machines to find which machine will give him/her maximum rewards.
* MABP is suitable for recommendation system as soloves the problem of recommedaing the content to produce maximum interactions(Rewards).
* MABP can be solved using different algorithms like **Epsilon greedy strategy, Upper confidence bound, Thompson Sampling**.
* Two goals are achieved using any of these methods,

1. Finding the correct arm that gives the maximum reward(arms are the posts)
2. Both Exploration and Exploitation can be used instead of using either one completely as that does not give maximum rewards.

* Epsilon greedy strategy is to use a exploration rate which will decide if the agent is going to explore other arms or if agent is going exploit the arm which give more rewards.
* Epsilon greedy is better to use when we have less number of arms(that is less no. of posts)
* As we have more arms it is better to use Thompson Sampling or Upper confidence bound.
* Thompson sampling solves the exploration-exploitation problem by using probablity of each arm.
* Upper confidence bound as the name says chooses the arm that has the upper bound.
* Contextual multi armed bandit is an enhanced method of solving the MABP.Where apart from the arms there is a state given which contains certain context on which the agent takes the action.
* Examples of context are :
  + - * + Age of the user
        + Gender
        + Demography of the user
* These contexts which are provided by the user can used as a criteria before recommending posts.
* Contextual bandit method can be integrated with Thompson Sampling using Bayesion linear regression/ neural network.