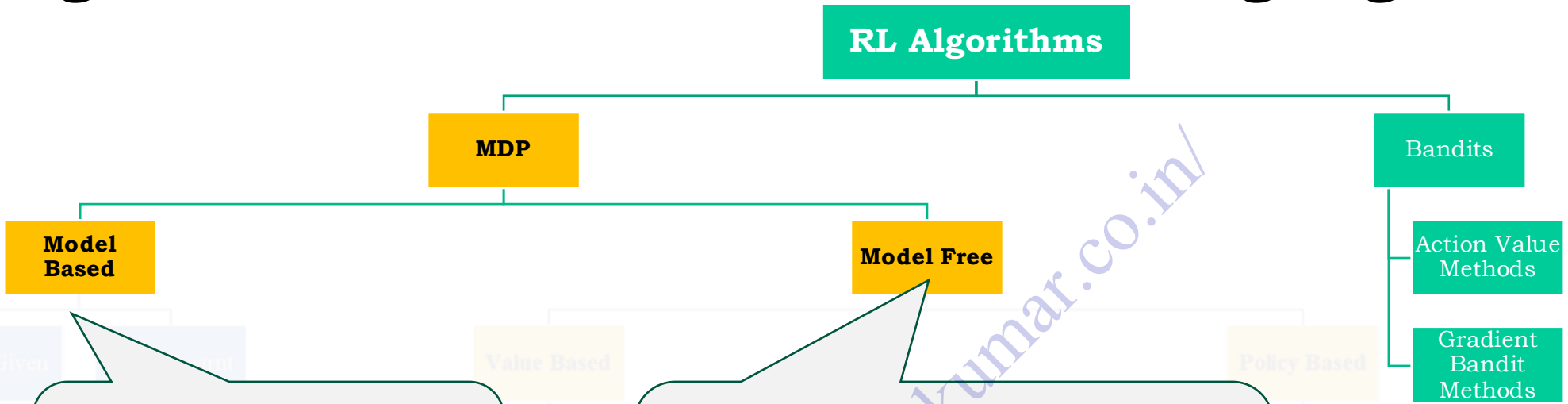
# Reinforcement Learning Algorithms

SARSA | Q – Learning | DQN | Double – DQN | PPO | TRPO | Policy Gradient | Actor – Critic | DDPG | A2C | A3C | SAC | TD3 | SAC

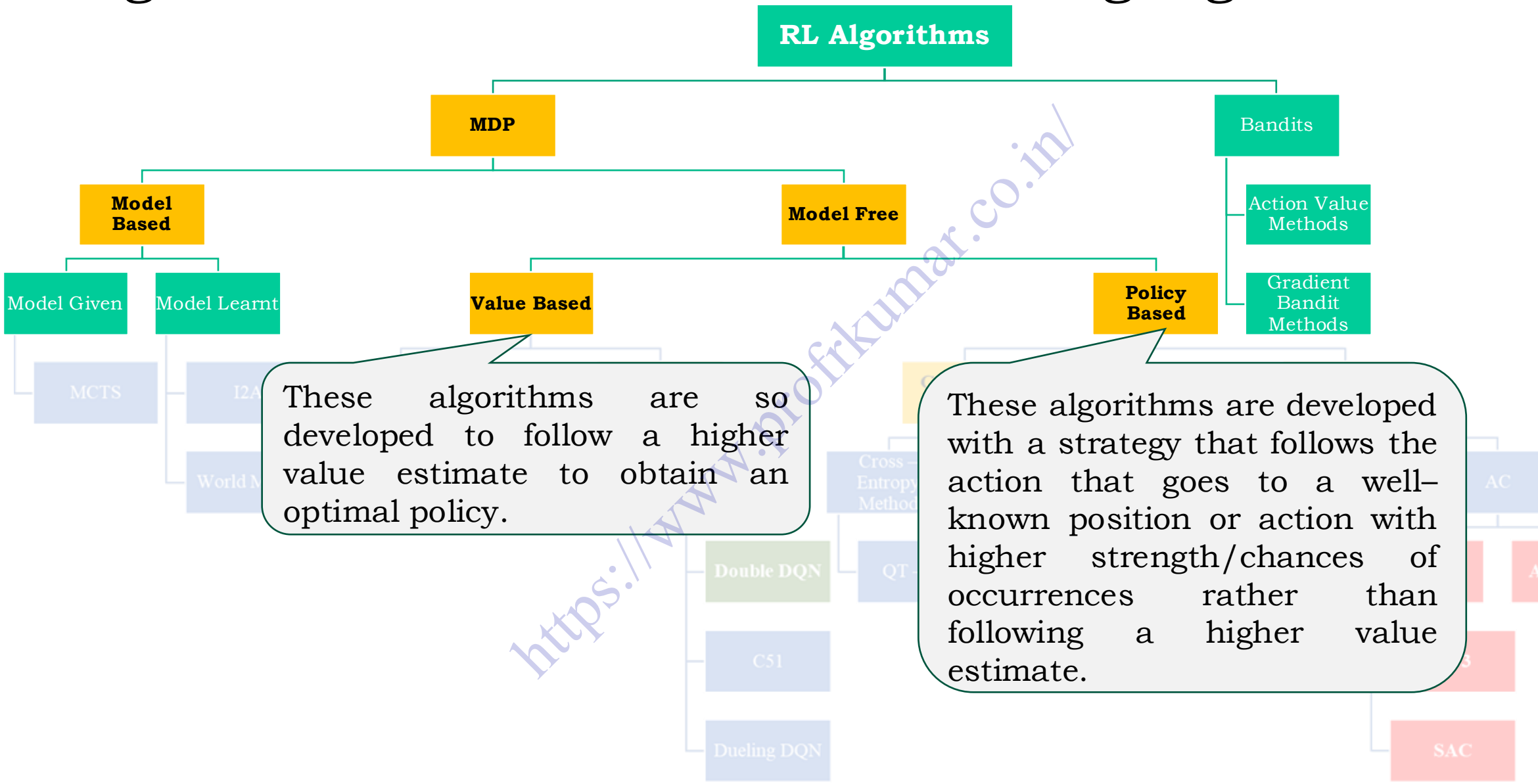
https://www.profrkumar.co.in/

# Categorization of Reinforcement Learning Algorithms



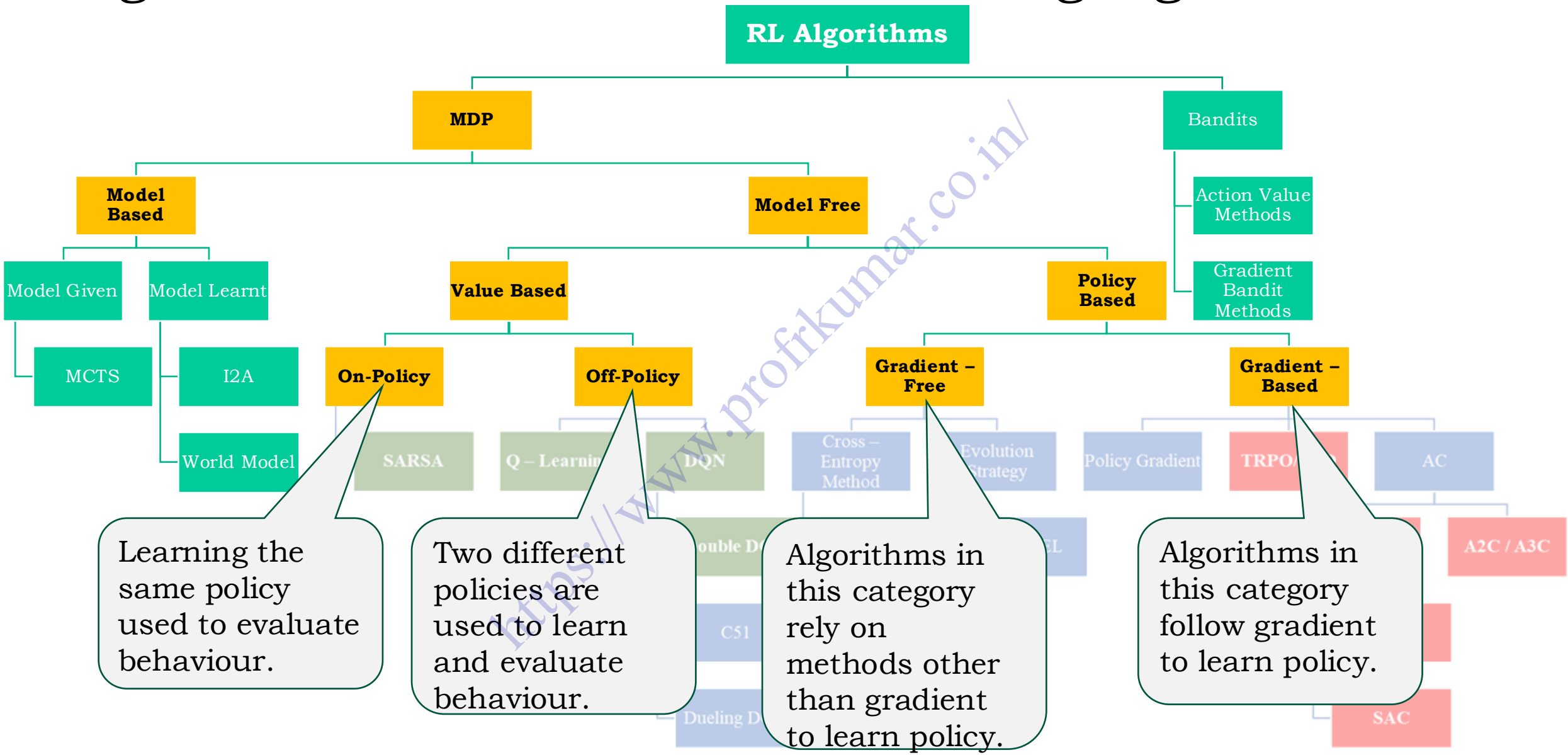

https://www.pradkumar.co.in/

# Categorization of Reinforcement Learning Algorithms



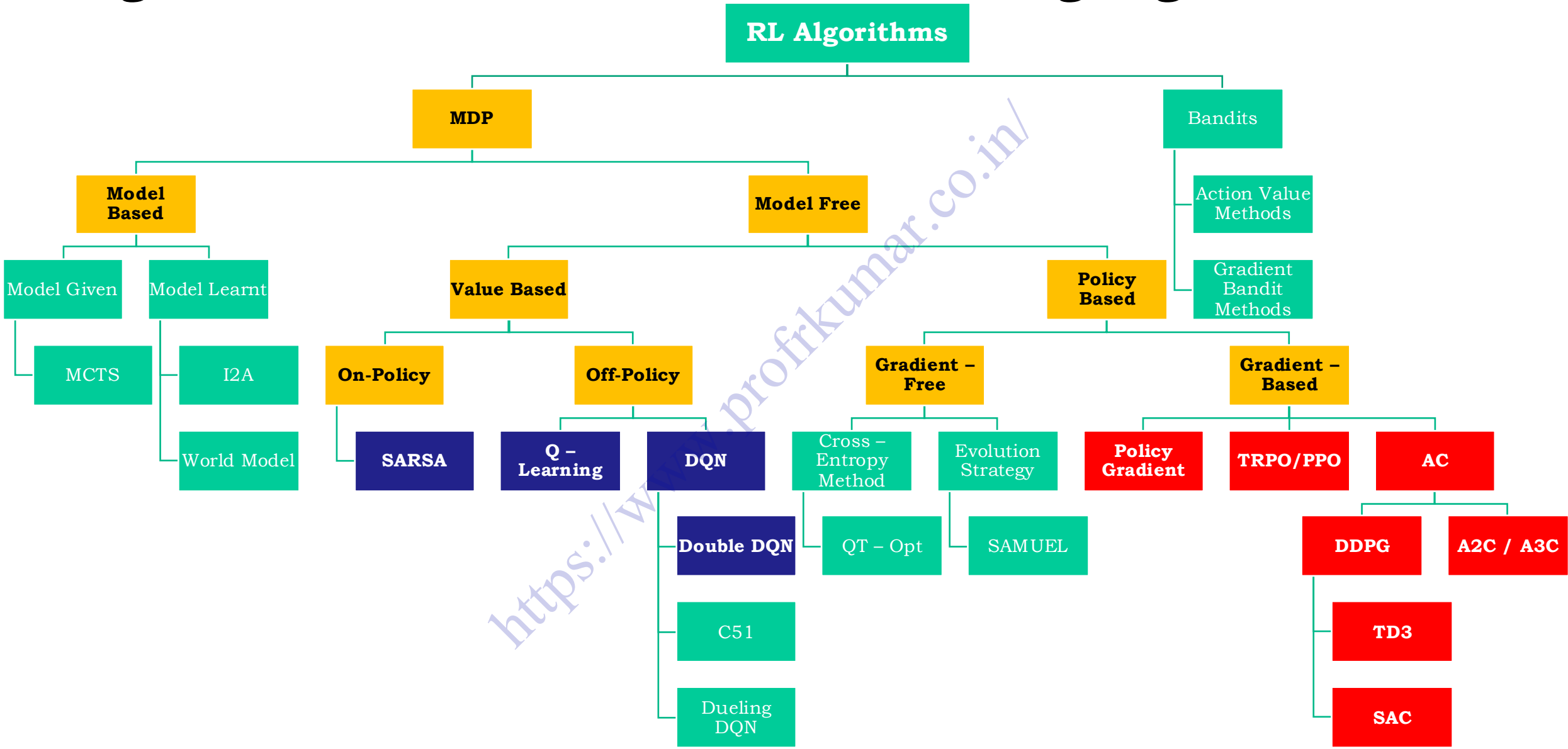

https://www.profrkumar.co.in/

# Categorization of Reinforcement Learning Algorithms




https://www.profrkumar.co.in/

# Categorization of Reinforcement Learning Algorithms




https://www.profrkumar.co.in/

# Timeline of key Reinforcement Learning (RL) algorithms

**Before 1980s – Foundations**

- **1952** – *Turing's reinforcement learning idea* in a theoretical sense.
- **1954** – *Farley & Clark*: First trial-and-error learning machine.
- **1959** – *Samuel's Checkers Program* (self-play with RL elements).
- **1960s-70s** – Early dynamic programming work by **Richard Bellman**

**1980s – Formalization of RL**

- **1983** – **Temporal Difference (TD) Learning** – *Sutton*
- **1989** – **Q-Learning** – *Watkins*

Model-free value iteration algorithm.

**1990s – Policy Search & Actor-Critic**

- **1992** – **Actor-Critic** Framework formalized.
- **1994** – **REINFORCE Algorithm** – *Williams*

First Monte Carlo policy gradient method.

- **1995** – **Eligibility Traces** – *TD(λ), Sarsa(λ).*
- **1998** – **Sarsa** Algorithm introduced.

https://www.profrkumar.co.in/

**2000s – Function Approximation & Robotics**

- **2000** – **Least-Squares Policy Iteration (LSPI)** – *Lagoudakis & Parr.*
- **2004** – **Fitted Q-Iteration** for continuous domains.
- **2005** – Use of **Gaussian Processes** in RL.
- **2008** – **Policy Gradient with Parameter-Based Exploration**.

**2013–2015 – Deep Reinforcement Learning (Deep RL)**

- **2013** – **Deep Q-Network (DQN)** – *Mnih et al., DeepMind,* Combines Q-learning.
- **2015** – **Double DQN** – Fixes Q-overestimation issue in DQN.
- **2015** – **Dueling DQN** – Separates state-value and advantage functions.

**2016–2018 – Actor-Critic and Continuous Control**

- **2016** – **Asynchronous Advantage Actor-Critic (A3C)** – *DeepMind,* Parallel agents learning.
- **2016** – **Deep Deterministic Policy Gradient (DDPG)** – *Lillicrap et al.*
- **2017** – **Proximal Policy Optimization (PPO)** – *OpenAI.*
- **2017** – **Trust Region Policy Optimization (TRPO)** – *Schulman et al.*
- **2018** – **Soft Actor-Critic (SAC)** – Entropy-regularized continuous control.
- **2018** – **TD3 (Twin Delayed DDPG)** – Addresses DDPG overestimation.

https://www.profrkumar.co.in/

**2019–2024 – Generalization, Sample Efficiency, Model-Based RL**

- **2019** – **Muzero** – *DeepMind*, Mastered games.
- **2020** – **DreamerV2** – World Models + Latent-space planning.
- **2021** – **Decision Transformer (DT)** – Reformulates RL as sequence modeling.
- **2022** – **Offline RL Boost** – Algorithms like Conservative Q-Learning (CQL).
- **2023** – **Gato** (DeepMind) – Generalist agent using RL + multitask learning.

**2025 and beyond – Trends**

- **Agentic AI** – Long-horizon planning, memory, and goal conditioning.
- **Foundation RL models** – Trained across tasks (RL + LLM hybrids).
- **Neuro-symbolic RL** – Combining logic, structure, and learning.
- **Energy-efficient RL** – RL on edge devices (neuromorphic + SNN).

https://www.profrkumar.co.in/