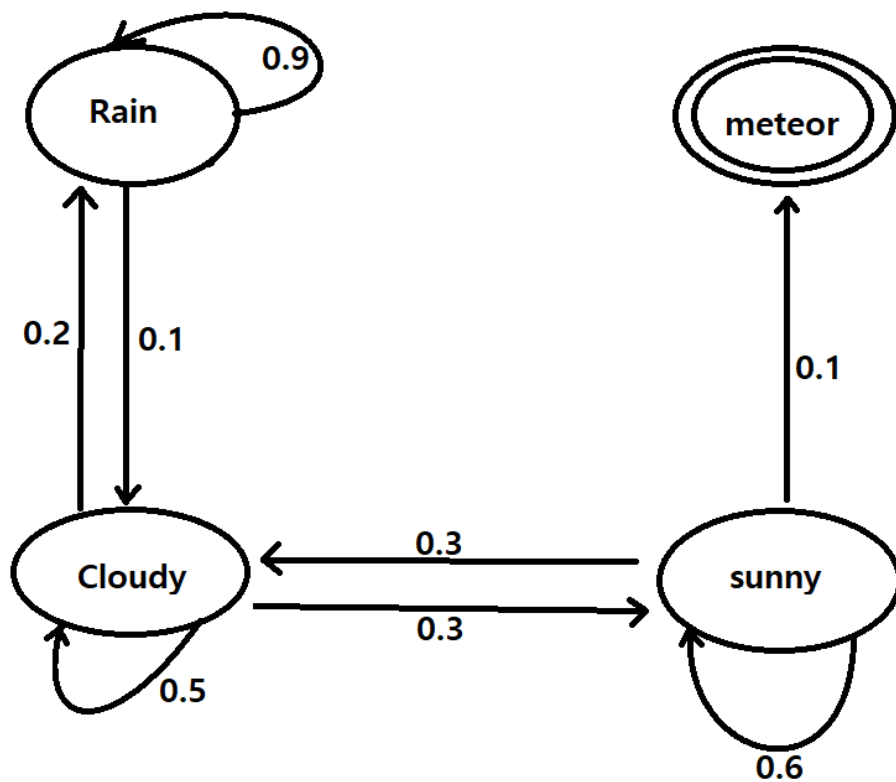


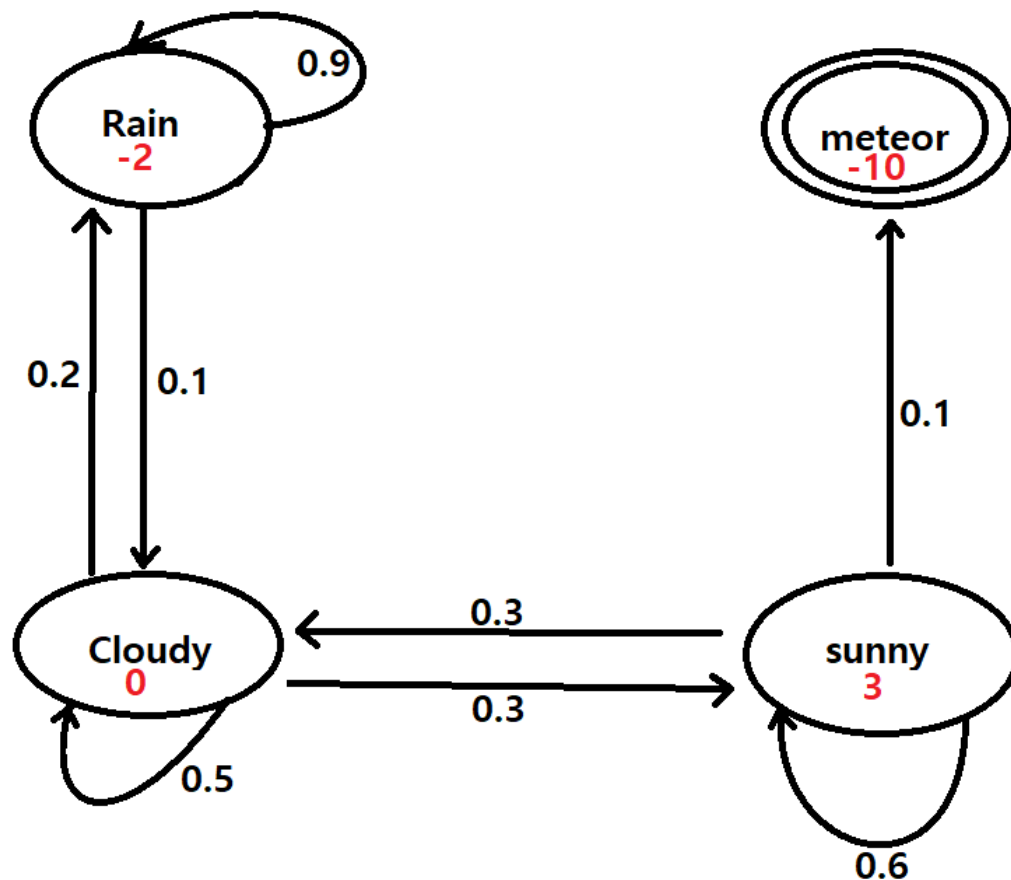
# Model-based Prediction and Control

## 1.1 Markov Chain

	Rain	cloudy	sunny	meteor
Rain	0.9	0.1	0.0	0
Cloudy	0.2	0.5	0.3	0
Sunny	0.0	0.3	0.6	0.1
meteor	0	0	0	0



## 1.2 Markov Reward Process



## 1.3 Sampling. Een voorbereiding voor Monte-Carlo Policy Evaluation

Discount = 1

Rain -> Rain -> Cloudy -> sunny -> meteor

$$0 + (1^1 * -2) + (1^2 * 0) + (1^3 * 3) + (1^4 * -10) = -9$$

sunny -> sunny -> sunny -> sunny -> meteor

$$0 + (1^1 * 3) + (1^2 * 3) + (1^3 * 3) + (1^4 * -10) = -1$$

## 1.4 De value-function bepalen

$$\gamma=1 \gamma=1$$

$$V(s) = E[G_t | S_t = s]$$

Iteration	Rain	Cloudy	Sunny	Meteor
0	0	0	0	0
1	-1.8	0.5	0.8	0
2	-3.37	0.63	1.43	0

### 1.5 Zelf-onderzoek

- De reward is onbegrensd en wordt daardoor eindeloos. [1]
- Het zorgt ervoor dat er geen focus gelegd kan worden op long-term en of short-term goals. [1]

## 2. Control met Value Iteration

Iteration	L	M	R
0	0	0	0
1	-0.1	-0.55	0
2	-0.65	-1	0
3	-1.1	-1	0
4	-1.1	-1	0
5	-1.1	-1	0

Ik ben gestopt met value iteration omdat er geen verandering meer is.

# Bronnen

1. <https://intuitivetutorial.com/2020/11/15/discount-factor/>