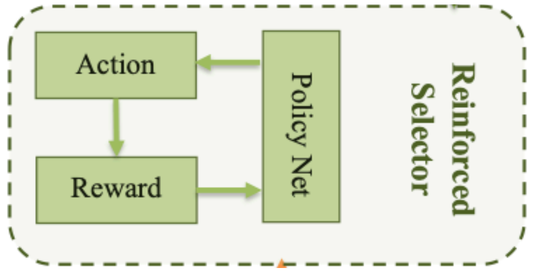
# Methodology

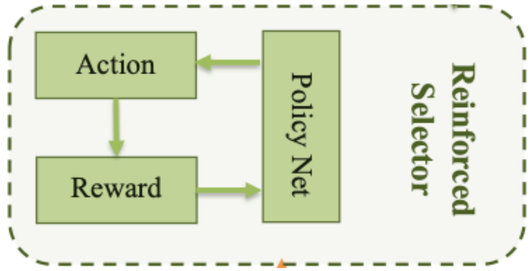
## Data Selection via Reinforcement Learning - *Action*




- The action value $a_i^{(k)}$ for every sample is 1 or 0.
  - 1: the action to *retain* the sample
  - 0: the action to *remove* the sample
- To determine the action, train a policy network includes two fully connected layers with corresponding activation functions, denote as $P\left(\cdot;\theta_s\right)$, $\theta_s$: the parameters
  - $P\left(s_i^{(k)};\theta_s\right)=\delta\left(\mathbf{w}_{s2}\cdot\mathrm{ReLU}\left(\mathbf{w}_{s1}\cdot s_i^{(k)}\right)\right)$
  - $\mathbf{w}_{s1},\mathbf{w}_{s2}$: weights of fully-connected layer, $\delta$: sigmoid activation function

# Methodology

## Data Selection via Reinforcement Learning - *Action*




- Then the action $a_i^{(k)}$ is sampled according to the output probability.
- The policy can be represented as:
- $$\pi_{\theta_s}\left(s_i^{(k)}, a_i^{(k)}\right) = \begin{cases} p_i^{(k)} & \text{if } a_i^{(k)} = 1 \\ 1 - p_i^{(k)} & \text{if } a_i^{(k)} = 0 \end{cases}$$