

# Podstawy Gymnasium

Bartosz Zieliński, Serhii Zeliuk  
Projekt 3

May 6, 2025

## 1. Wprowadzenie

Celem projektu było zapoznanie się z podstawami uczenia ze wzmocnieniem przy użyciu środowisk gymnasium. Projekt został zrealizowany w trzech częściach: rozwiązanie klasycznego problemu Taxi przy użyciu Q-learningu, rozwiązanie problemu z przestrzenią ciągłą (MountainCarContinuous-v0), a także implementacja własnego środowiska i agenta do gry Blackjack, z porównaniem kilku metod.

## 2. Zadanie 1: Taxi-v3

### 2.1 Opis środowiska

Taxi-v3 to klasyczne środowisko ze skończoną liczbą stanów i akcji, gdzie agent (taksówka) ma za zadanie odebrać i odstawić pasażera do celu. Każdy ruch kosztuje -1 punkt, błędne podniesienie pasażera -10, a sukces to +20 punktów.

### 2.2 Parametry uczenia

- Algorytm: Q-learning
- Liczba epizodów: 1000
- Współczynnik uczenia:  $\alpha = 0,1$
- Współczynnik dyskontowy:  $\gamma = 0,9$
- Początkowa eksploracja:  $\varepsilon = 1.0$ , malejąca do 0.1

### 2.3 Wyniki

Najlepsza suma nagród osiągnięta przez agenta wyniosła 15 punktów w 864 epizodzie. Wyniki pokazują, że agent z czasem nauczył się podejmować lepsze decyzje.

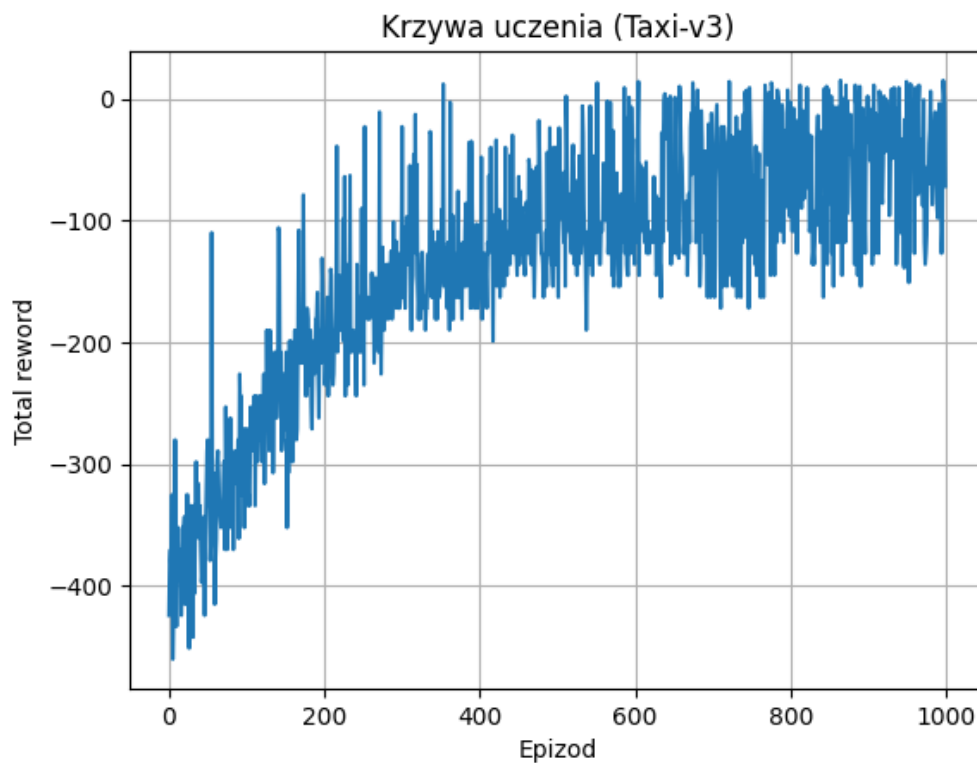


Figure 1: Krzywa uczenia agenta Taxi (suma nagród na epizod)

### 3. Zadanie 2: MountainCarContinuous-v0

#### 3.1 Przestrzeń ciągła

W tym środowisku agent musi rozpedzić się pod górke, korzystając z przestrzeni ciągłej. Problem został zdyskretyzowany dla potrzeb Q-learningu.

#### 3.2 Eksperymenty z AI

W ramach eksperymentu porównaliśmy wpływ trzech różnych wartości  $\gamma$ :

- $\gamma = 0,8$
- $\gamma = 0,9$
- $\gamma = 0,99$

Wyniki pokazały, że większy  $\gamma$  (czyli większe znaczenie przyszłych nagród) poprawia zdolność agenta do nauki.

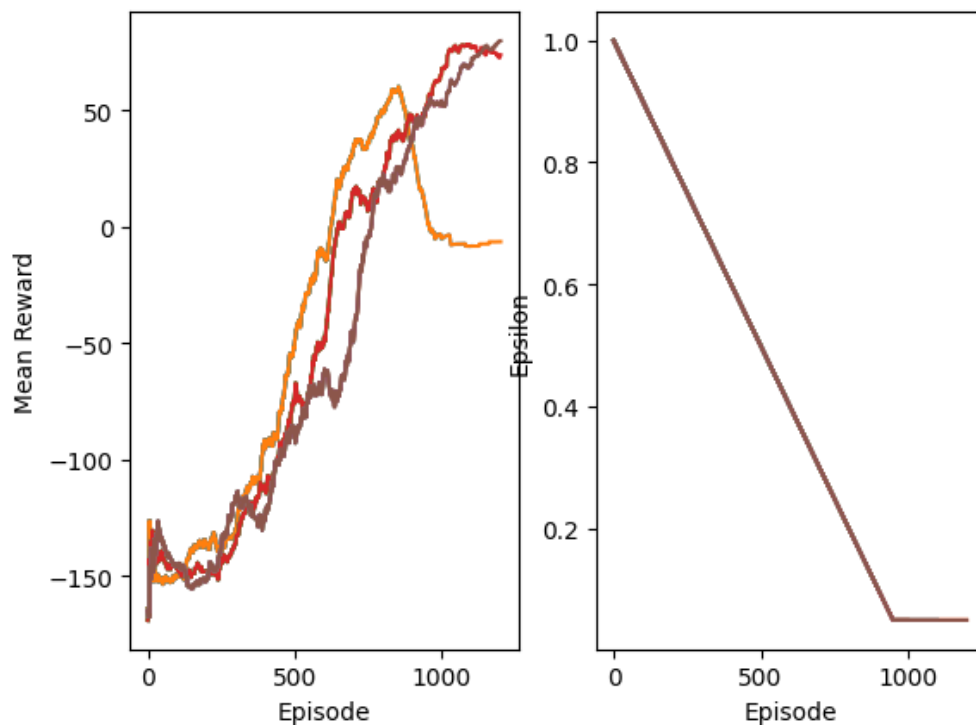


Figure 2: Krzywe uczenia agenta MountainCarContinuous dla różnych  $\gamma$

## 4. Zadanie 3: Blackjack

### 4.1 Algorytmy

W tym zadaniu zaimplementowano trzech agentów:

- Q-learning
- Monte Carlo
- Q-learning z oceną ryzyka (risk-aware)
- Dodatkowo: DQN z siecią neuronową i licznikiem kart

### 4.2 Wyniki

Wyniki sugerują, że podejście *risk-aware* oraz DQN dawały lepsze wyniki niż klasyczny Q-learning. Współczynnik zwycięstw wzrósł do ponad 44

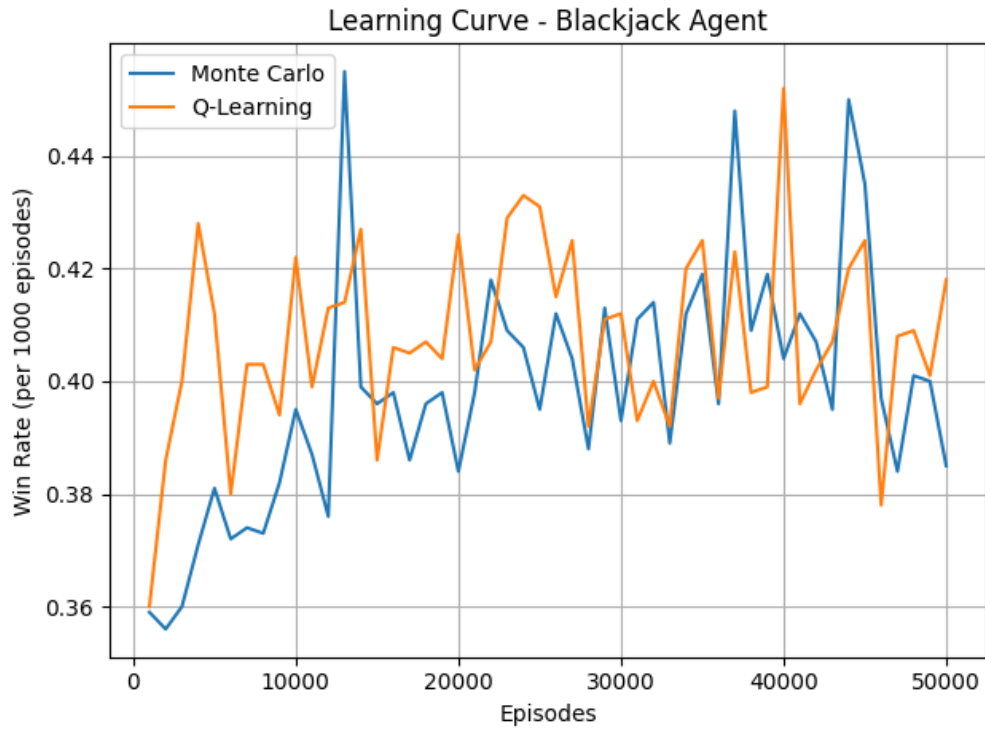


Figure 3: LC dla podejścia bez brania pod uwagę ryzyka

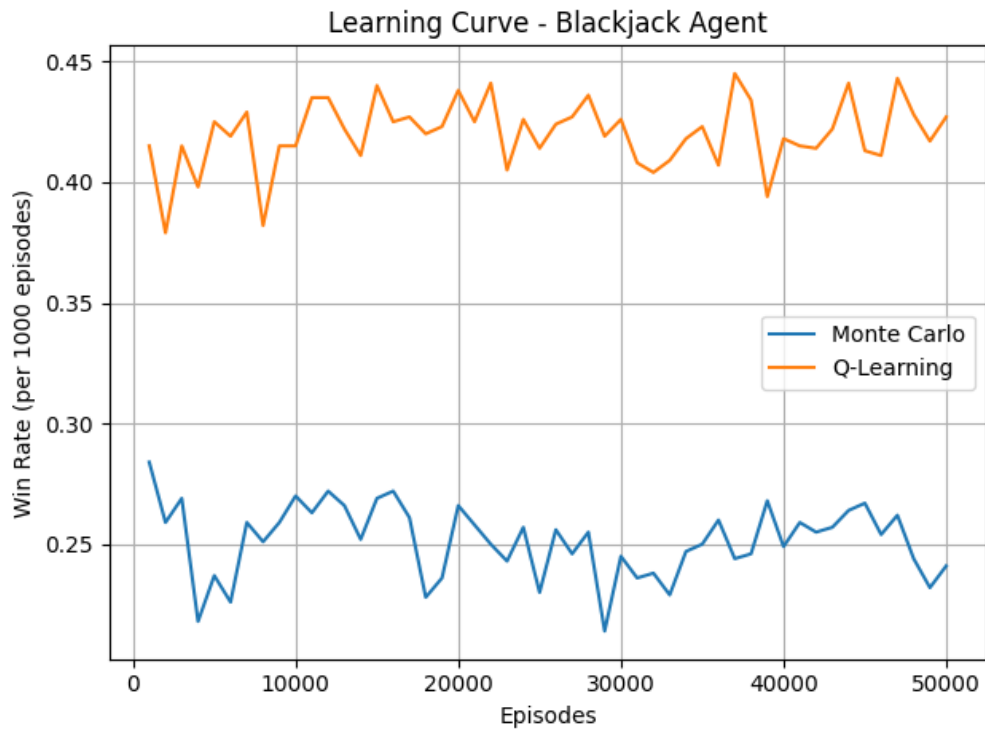


Figure 4: LC biorąc pod uwagę zliczanie kart

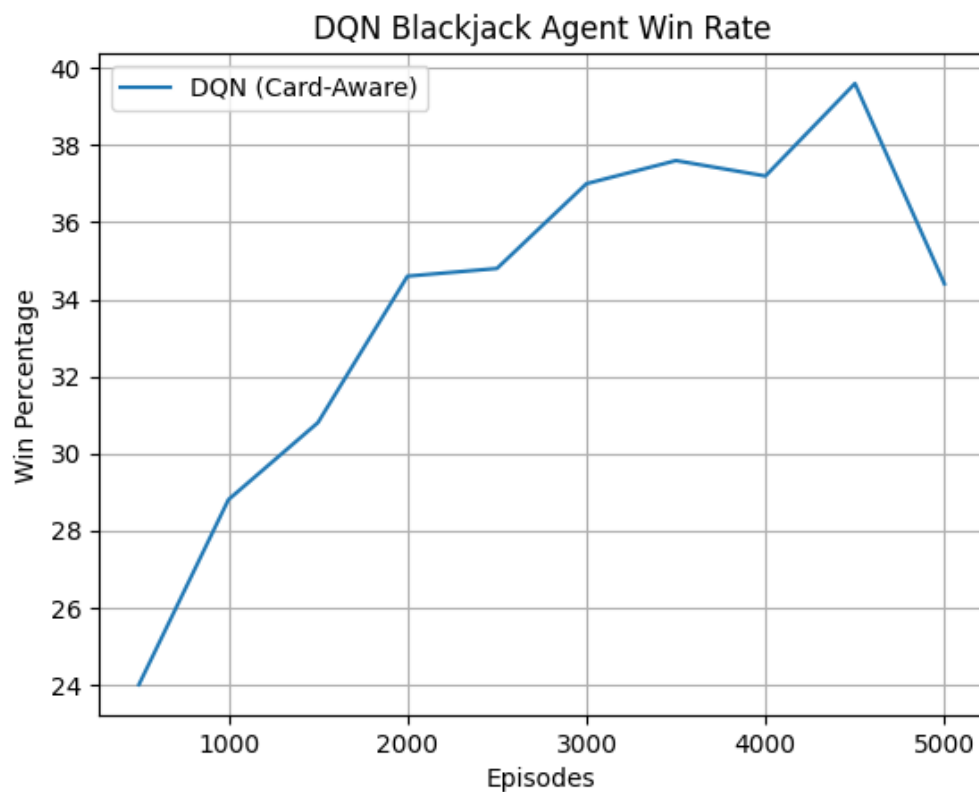


Figure 5: LC dla sieci neuronowej (DQN)

## 5. Podsumowanie

Projekt umożliwił praktyczne poznanie różnych metod uczenia ze wzmocnieniem w środowiskach dyskretnych i ciągłych. Przeprowadziliśmy eksperymenty z parametrami  $\gamma$ ,  $\varepsilon$  i  $\alpha$ , a także porównaliśmy klasyczne i nowoczesne podejścia (Monte Carlo, DQN).