МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

им. Н.Э. Баумана

Факультет «Информатика и системы управления»

Кафедра «Систем обработки информации и управления»

ОТЧЕТ

**Лабораторная работа №\_\_4\_**

по дисциплине«Методы машинного обучения в автоматизированных системах обработки информации и управления»

Тема: « Реализация алгоритма Policy Iteration»

ИСПОЛНИТЕЛЬ: \_\_Се Цзявэнь\_\_\_\_

ФИО

группа ИУ5И-22М\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

подпись

"\_\_"\_\_\_\_\_\_\_\_\_202\_ г.

ПРЕПОДАВАТЕЛЬ: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

ФИО

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

подпись

"\_22\_"\_\_\_\_\_05\_\_\_\_2024\_ г.

Москва - 2024

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# Задания

* 1. На основе рассмотренного на лекции примера реализуйте алгоритм Policy Iteration для любой среды обучения с подкреплением (кроме рассмотренной на лекции среды Toy Text / Frozen Lake) из библиотеки Gym (или аналогичной библиотеки)..

# FrozenLake8x8:

The FrozenLake8x8 environment is a classic maze problem in OpenAI Gym, challenging players to find the end point through a frozen lake.

# метод:

Алгоритм итерации политики — это независимый от модели алгоритм обучения с подкреплением, используемый для поиска оптимальной политики. Он достигает своих целей, чередуя два этапа: оценку стратегии и улучшение стратегии.

1. Инициализация: сначала мы инициализируем случайную политику и функцию значения. Стохастическая политика означает, что вероятность выбора действия в каждом состоянии одинакова. Функция стоимости представляет собой ожидаемую прибыль, которую можно получить, приняв стратегию в каждом состоянии.
2. Оценка политики. На этапе оценки политики мы оцениваем функцию ценности каждого состояния на основе текущей политики. Мы используем уравнение Беллмана для обновления функции ценности, которая включает взвешенное суммирование всех возможных следующих состояний и вознаграждений.
3. Улучшение политики: на этапе улучшения политики мы улучшаем политику на основе текущей функции стоимости. Для каждого состояния мы пробуем все возможные действия и выбираем действие, которое может получить функцию максимального значения в качестве нового оптимального действия.

## выполнить

Мы используем язык программирования Python и библиотеку OpenAI Gym для реализации алгоритма итерации политики для решения проблемы среды FrozenLake8x8.

Сначала импортируем необходимые библиотеки и среду FrozenLake8x8. Затем мы определяем функцию итерации политики policy\_iteration(), которая принимает в качестве входных данных среду, коэффициент дисконтирования гамму и максимальное количество итераций max\_iterations.

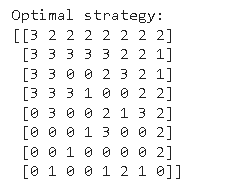
В функции policy\_iteration() мы инициализируем случайную функцию политики и значения. Затем мы начинаем итерации оценки и улучшения стратегии.

На этапе оценки политики мы используем уравнение Беллмана для обновления функции стоимости до тех пор, пока функция стоимости не сойдется. На этапе улучшения политики мы выбираем новое оптимальное действие на основе текущей функции стоимости.

Наконец, когда политика стабилизируется или достигает максимального количества итераций, мы возвращаем оптимальную политику и функцию ценности.

Найдите все выбросы и замените их медианой

Согласно алгоритму итерации политики оптимальная стратегия находится в среде FrozenLake8x8. Оптимальная стратегия — это матрица 8x8, представляющая оптимальное действие, которое следует предпринять в каждой позиции. 



# Вывод

Используя алгоритм итерации политики, мы успешно решили среду FrozenLake8x8 и нашли оптимальную политику. Алгоритм итерации политики — это эффективный алгоритм обучения с подкреплением, который может найти оптимальную политику в среде, независимой от модели..