Vietnam General Confederation of Labor

**TON DUC THANG UNIVERSITY**

**FACULTY OF INFORMATION TECHNOLOGY**



**FINAL PROJECT**

**Course: Introduction to Artificial Intelligence**

*Instructor*:  **Msc. Nguyen Thanh An**

*Student*: **Tran Le Minh Tri – 520H0500**

*Year***: 24**

**HO CHI MINH CITY, 2024**

Vietnam General Confederation of Labor

**TON DUC THANG UNIVERSITY**

**FACULTY OF INFORMATION TECHNOLOGY**



**FINAL PROJECT**

**Course: Introduction to Artificial Intelligence**

*Instructor*:  **Msc. Nguyen Thanh An**

*Student*: **Tran Le Minh Tri– 520H0500**

*Class***: 20H50303**

*Year***: 24**

**HO CHI MINH CITY, 2024**

ACKNOWLEDGEMENT

The first sincere thanks I want to give to **Mr. Nguyen Thanh An**, who enthusiastically taught and worked tirelessly to give me enough tools and skills to complete this report. She played an important role in improving my mathematical logic and knowledge. The second thanks I would like to give to the teachers of the Department of Information Technology of Ton Duc Thang University for giving me the opportunity to do this report, because it is not only a report but also a very important experience for me in the next 3 years.

My report will have some errors, I am very open to receiving feedback from teachers so that I can improve my report writing skills.

Finally, I wish you good health and success in your noble career.

*Ho Chi Minh city, 19th May, 2022*

*Author*

*(Sign and write full name)*

*Tran Le Minh Tri*

THIS PROJECT WAS COMPLETED AT

TON DUC THANG UNIVERSIY

I fully declare that this is my own project and is guided by **Mr**. **Nguyen Thanh An**. The research contents and results in this topic are honest and have not been published in any form before. The data in the tables for analysis, comments and evaluation are collected by the author himself from different sources, clearly stated in the reference section.

Besides that, the project also uses a number of comments, assessments as well as data from other authors, other agencies and organizations, with citations and source annotations.

**Should any frauds be found, I will take full responsibility for the content of my report.** Ton Duc Thang University is not related to copyright and copyright violations caused by me during the implementation process (if any).

*Ho Chi Minh city, 19th May, 2024*

*Author*

*(Sign and write full name)*

*Tran Le Minh Tri*

CONFIRMATION AND ASSESSMENT SECTION

**Instructor confirmation section**

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

*Ho Chi Minh city, 19th May, 2024*

*(Sign and write full name)*

**Evaluation section for grading instructor**

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

*Ho Chi Minh city,19th May, 2024*

*(Sign and write full name)*

Contents

[Chapter 1. Student List 6](#_Toc167014253)

[Chapter 2. Solve Task 1 6](#_Toc167014254)

[2.1. Synthesize classes in code 6](#_Toc167014255)

[2.2. Pseudocode Tic Tac Toe 7](#_Toc167014256)

[2.3. How to play Tic Tac Toe 7](#_Toc167014257)

[2.4. Result of Tic Tac Toe 9](#_Toc167014258)

[2.5. Practical examples 9](#_Toc167014259)

[2.6. Advantages versus disadvantages 10](#_Toc167014260)

[Chapter 3. Solve Task 2 10](#_Toc167014261)

[3.1 Introduction 10](#_Toc167014262)

[3.1.1 What is n-queen problem and why is it important? 10](#_Toc167014263)

[3.1.2 What is CNFs? 10](#_Toc167014264)

[3.2 Pseudocode 11](#_Toc167014265)

[3.3 Result 12](#_Toc167014266)

[3.3 Advantage and Disadvantage 13](#_Toc167014267)

[3.3.1 Advantage: 13](#_Toc167014268)

[3.3.2 Disadvantage: 13](#_Toc167014269)

[Chapter 4. Solve Task 3 14](#_Toc167014270)

[4.1 Introduction 14](#_Toc167014271)

[4.2 Calculating Entropy, Average Entropy, and Information Gain 14](#_Toc167014272)

[4.3 Implementing, Training, and Evaluating a Decision Tree Model 15](#_Toc167014273)

[4.4 Advantages and disadvantages 15](#_Toc167014274)

# Chapter 1. Student List

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Student ID** | **Full Name** | **Email** | **Assigned Tasks** | **Complete Percentage** |
| 520H0639 | Trương Thái Gia Hưng | 520H0639@student.tdtu.edu.vn | 3 | 100% |
| 522H0148 | Phạm Đặng Thanh Trung | 522H0148@student.tdtu.edu.vn | 2 | 100% |
| 522H0142 | Nguyễn Thành Long | 522H0142@student.tdtu.edu.vn | 1 | 100% |
| 520H0500 | Trần Lê Minh Trí | 520H0500@student.tdtu.edu.vn | 3 | 100% |

# Chapter 2. Solve Task 1

## 2.1. Synthesize classes in code

* Class Problem
* \_\_init\_\_: Create matrix 2D 8x8.
* draw: Print column numbers, horizontal divider, row numbers and board state
* make move: Update the game board with the player's move at the specified row and column.
* is\_valid\_move: Check if the move is valid.
* check\_winner: Check winner by check\_win\_direction.
* check\_win\_direction: Make direction to win.
* is\_full: Check the board is full.
* can\_create\_four\_in\_a\_row: Check out the possible winning lines.
* can\_block\_four\_in\_row: Find and block your opponent’s winning lines.
* get\_action: Returns a list of actions that can be taken from the current state of the problem.
* result: Performs an action from the current state of the problem.
* is\_terminal: Checks whether the game has ended.
* utility: Calculates the utility function for the current state of the problem.
* Class SearchStrategy
* alpha\_beta\_search, abminimax: Implements the alpha-beta pruning algorithm.
* Class Game
* main: Play game tic tac toe.

## 2.2. Pseudocode Tic Tac Toe

**function** ALPHA-BETA-SEARCH (state) **returns** an action

v ← MAX-VALUE (state, -∞, +∞)

**return** the action in ACTIONS (state) with value v

**function** MAX-VALUE (state, α, β) **returns** a utility value

**if** TERMINAL-TEST (state) **then** **return** UTILITY (state)

v ← -∞

**for** **each** a **in** ACTIONS (state) **do**

v ← MAX (v, MIN-VALUE (RESULT (s, a), α, β))

**if** v ≥ β **then** **return** v

α ← MAX (α, v)

**return** v

**function** MIN-VALUE (state, α, β) **returns** a utility value

**if** TERMINAL-TEST (state) **then** **return** UTILITY (state)

v ← +∞

**for** **each** a **in** ACTIONS (state) **do**

v ← MIN (v, MAX-VALUE (RESULT (s, a), α, β))

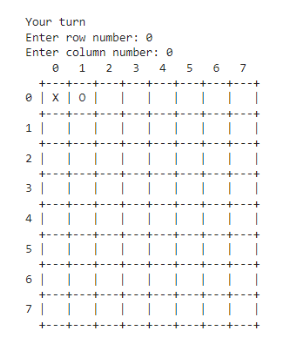
**if** v ≤ α **then** **return** v

β ← MIN (β, v)

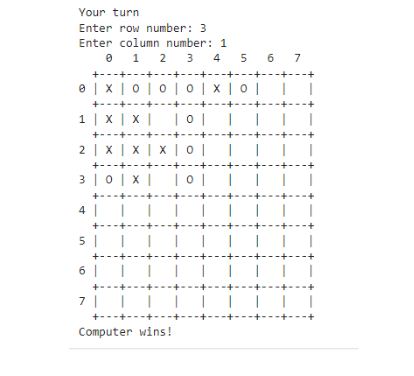
**return** v

## 2.3. How to play Tic Tac Toe

1. *You are X. Computer is O.*
2. *When the screen displays “your turn”.*
3. *Enter the number of row you want.*
4. *Enter the number of column you want.*



## 2.4. Result of Tic Tac Toe



## 2.5. Practical examples

* The alpha beta pruning algorithm is applied in practical examples:
* Computer games: tic tac toe, chess….
* Information retrieval
* Robot planning and decision-making systems
* Dynamic programming optimization

## 2.6. Advantages versus disadvantages

* Advantages:
* Improved Efficiency
* Better Decision Making
* Memory Savings
* Broad Applicability
* Disadvantages:
* Potential for Suboptimal Decisions
* Sensitivity to Move Ordering
* Limited Applicability to Non-Adversarial Problems
* Complexity Analysis

# Chapter 3. Solve Task 2

## 3.1 Introduction

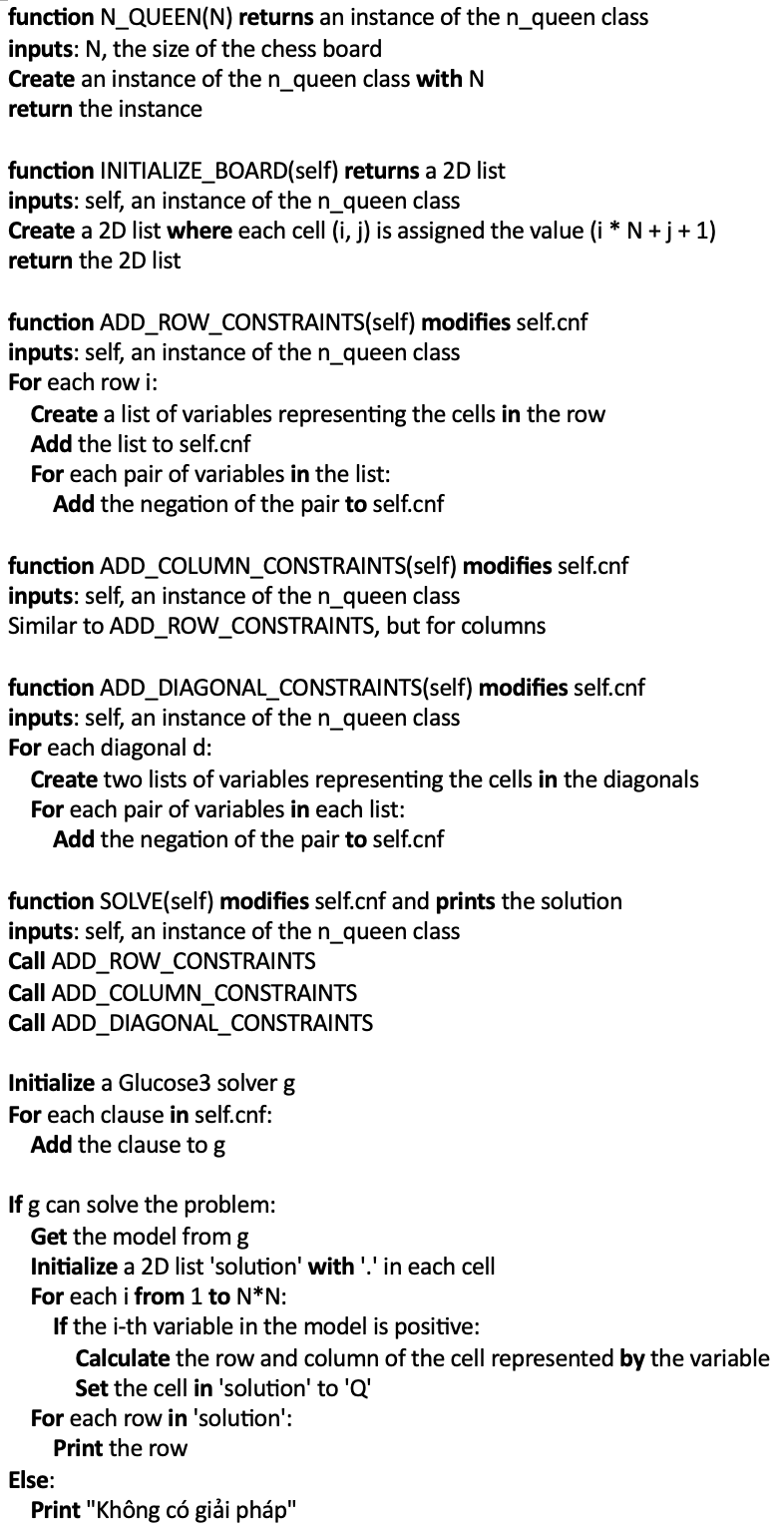
### 3.1.1 What is n-queen problem and why is it important?

The N-Queens problem is a problem in the field of artificial intelligence (AI) and computer science, in which the goal is to place N queens on an N x N chess board such that no two queens threaten threaten each other. This means that no two queens are on the same row, column or diagonal. This problem is an example of a constraint problem and is widely used in AI research to test and compare constraint search and solving algorithms.

### 3.1.2 What is CNFs?

* CNF (Conjunctive Normal Form) is a standard form of a logical expression in which the expression is written as a union (AND) of one or more clauses, each clause is an union (OR) of one or more expressions.
* In the N-Queens problem, we use CNF to model the constraints of the problem. Each queen on the chessboard is represented by a variable. Each constraint (no two queens should be in the same row, column or diagonal) is represented by a clause in CNF.
* Using CNF makes it possible to use SAT solvers (like Glucose3 in this case) to solve the problem. These tools have been optimized to solve SAT problems efficiently, helping to solve the N-Queens problem more quickly than using traditional search or backtracking methods.

## 3.2 Pseudocode



**\_\_init\_\_(self, N)** -> Initialize the class with checkerboard size N x N, an empty list to store the CNF conditions, and an initialized table.

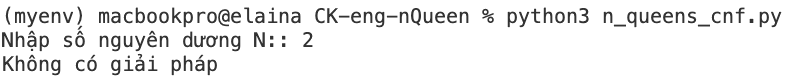
**initialize\_board(self)** : -> Initialize the board with numbers from 1 to N^2.

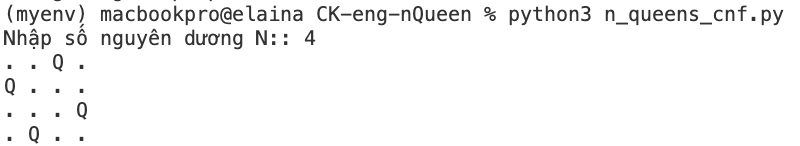
**add\_row\_constraints(self), add\_column\_constraints(self) and add\_diagonal\_constraints(self)**: -> Add constraints to the rows, columns and diagonals of the table so that no two queens are on the same row, column or diagonal.

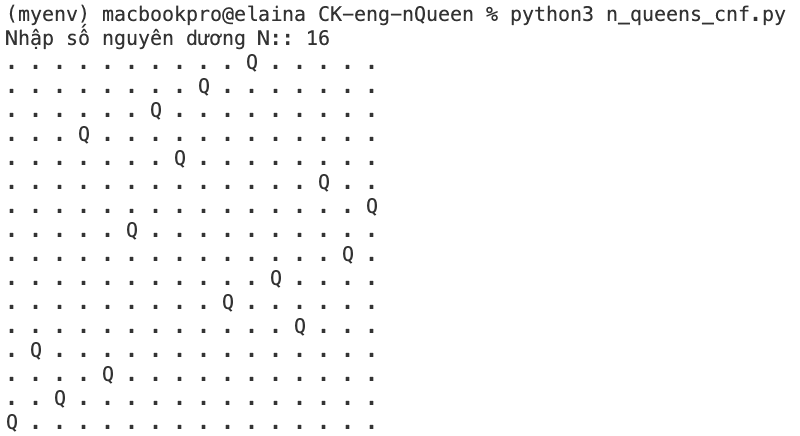
**solve(self):** ->Use SAT solver (Glucose3) to solve the problem. If there is a solution, it prints the table with the positions of the queens. Otherwise, it will print the message "Không có giải pháp".

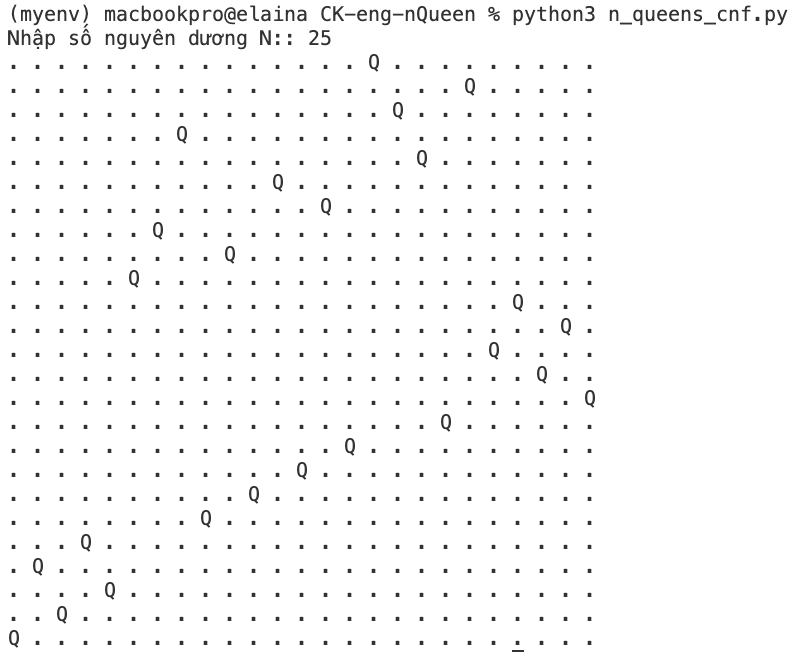
## 3.3 Result











## 3.3 Advantage and Disadvantage

### 3.3.1 Advantage:

* This method can be applied to any size of chess board, not just limited to 8x8 chess board.
* Systematic properties: Using CNFs allows us to model the problem systematically and logically, making it easier to control and better understand the problem.
* Using SAT solver: Using SAT solver like Glucose3 helps take advantage of the power of modern SAT solver algorithms, which can solve problems quickly and effectively.

### 3.3.2 Disadvantage:

* Complexity: Generating CNFs for the N-Queens problem can generate a large number of logic expressions, especially as the chessboard size increases. This can increase the complexity of the problem.
* Scalability: Although this method can be applied to any chessboard size, as the chessboard size increases, the time to solve the problem also increases significantly.
* Knowledge requirements: To understand and effectively use this method, users need to have solid knowledge of propositional logic and SAT solver.

# Chapter 4. Solve Task 3

## Introduction

The objective of this report is to calculate the Entropy (H), Average Entropy (AE), and Information Gain (IG) for a specific score attribute (Q1 to Q9) and to implement, train, and evaluate a decision tree model using the given dataset. Finally, we will visualize the structure of the decision tree.

## Calculating Entropy, Average Entropy, and Information Gain

1. Load and Clean Data:

* Load the dataset and remove unnecessary columns.
* Check for and display any duplicate rows.
* Encode the Rank column to numerical values.

1. Entropy Calculation:

* **Entropy (H)**: Entropy measures the uncertainty or impurity in a dataset. The formula for entropy is:
* 

1. Average Entropy and Information Gain Calculation:

* **Average Entropy (AE)**: Average Entropy of an attribute with respect to the target variable



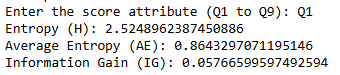
* **Information Gain (IG)**: Information Gain is the reduction in entropy after knowing the value of an attribute.



**Result:**

**Input:** Q1

**Output:**

****

## Implementing, Training, and Evaluating a Decision Tree Model

1. **Prepare the Data**:

* Define feature variables and target variables.
* Split data into training and testing sets.

1. **Train the Model**:

* Initialize and train the Decision Tree model.

1. **Evaluate the Model**:

* Make predictions on the test set.
* Calculate accuracy and other evaluation metrics.

1. **Visualize the Decision Tree**:

* Use visualization tools to display the structure of the decision tree.

## Advantages and disadvantages

Decision Trees:

* **Advantages**:

1. Easy to interpret and visualize.
2. Can handle both numerical and categorical data.
3. Requires little data preprocessing.

* **Disadvantages**:

1. Prone to overfitting, especially with complex trees.
2. Can be unstable with small variations in data.
3. Greedy algorithms may not produce the optimal tree.

Result:

