Nama : Shidqi Aqil Naufal

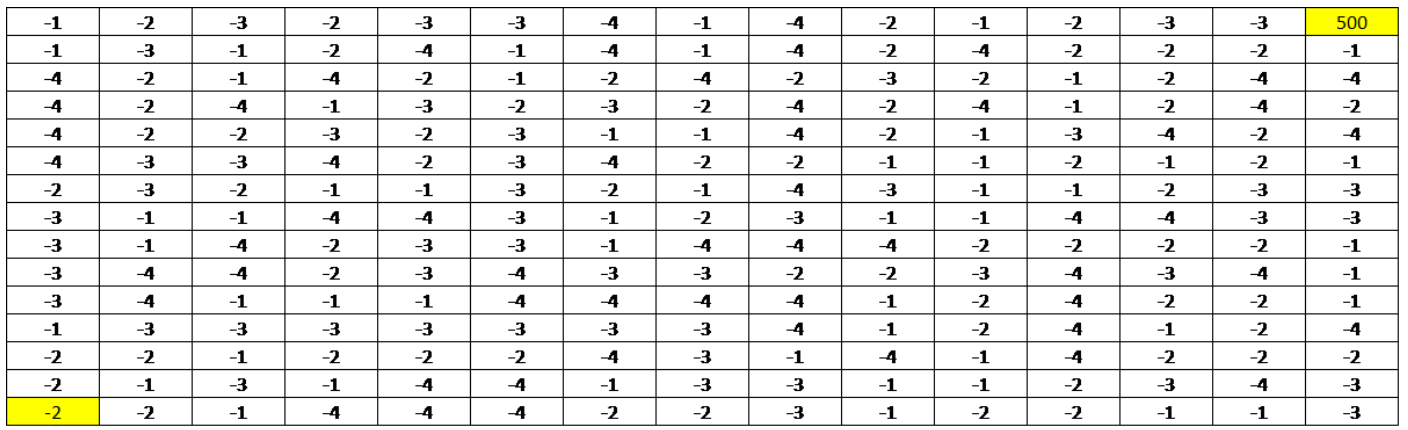
Nim : 1301164147

Kelas : IF40-08

**Laporan Tugas 3 Mencari Jalan Optimum dengan Metode Q-Learning**

**DESKRIPSI MASALAH**

Diberikan sebuah permasalahan untuk mencari suatu jalan ketujuan yang telah diberikan. Terdapat sebuah word grid yang telah diberikan beserta rewardnya disetiap kotaknya. Adapun gambaran word grid yang diberikan adalah sebagai berikut:



Terdapat dua tanda kuning apada word grid diatas. Pada word grid (1,1) adalah inisiasi atau langkah awal dimana kita akan memulai perjalanan. Pada titik awal (1,1) mempunyai reward -2. Tujuan atau garis finish dari grid diatas adalah pada grid (15,15) yang mempunyai reward 500. Adapun langkah yang dapat dilakukan adalah 4 arah, antara lain: north, west, east, dan south. Hanya 4 langkah ini yang dapat digunakan untuk mencapai tujuan yang ditentukan.

**METODE PENYELESAIAN**

Metode Q-Learning akan digunakan untuk menyelesaikan pencarian jalur diatas. Langkah langkah dalam metode Q-lEarning adalah sebagi berikut:

1. Perisapkan tabel Q(s,a), yang mana tabel ini berfungsi untuk menyimpan hasil dari setiap step yang dilakukan. Tabel Q(s,a) akan dibuat dalam bentuk sebagi berikut :

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **N** | **E** | **S** | **W** |
| **0** | 0 | 0 | 0 | 0 |
| **1** | 0 | 0 | 0 | 0 |
| **..** | 0 | 0 | 0 | 0 |
| **225** | 0 | 0 | 0 | 0 |

Semua isi tabel diinisiasi dengan 0 semua. Terdapat 225 langkah yang dipersiapkan untuk menampung setiap step yang dilakukan oleh sistem. Adapun langkah yang dipersiapkan terdapat 4 langkah antara lain, North, East, South, West. Tabel ini kan di update pada setiap step yang dilakukan.

1. Langkah selanjutnya adalah eksplorasi. Langkah ini bertujuan untuk mencari jalan hingga mencapai tujuan yang telah ditentukan. Proses eksplorasi terdiri dari 5000 episode yang mana setiap episodenya terdiri dari 100 step. Pada setiap step akan dilakukan update dari tabel Q(s,a). rumus dari Q(s,a) adalah sebagai berikut:

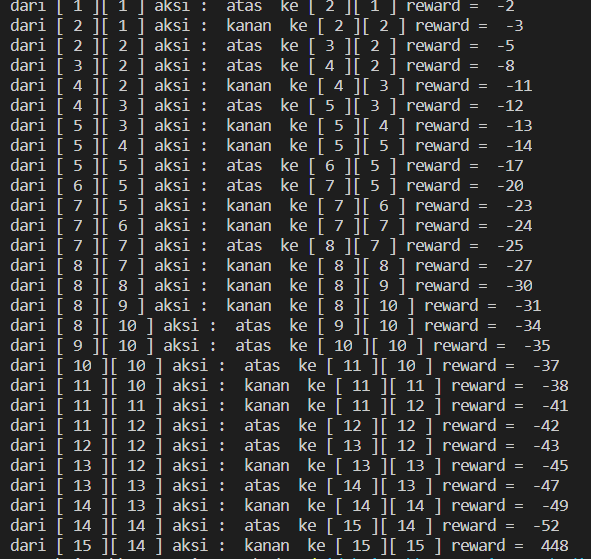
Q(s,a) = Q(s,a) + α[ r + ᵧmax Q(s’,a’) – Q(s,a) ]

α adalah alpha, yang mana kita dapat menentukan besarnya. Kemudiana r adalah reward, ᵧ max adalah maksimum pada tabel Q(s,a) pada setiap arahnya.

1. Setelah proses ekploitasi selesai, lakukan proses ekplorasi, yang mana proses ini akan memilih jalan yang mempunyai nilai paling besar disetiap langkahnya. Nilai yang diambil berdasarkan isi dari tabel Q(s,a). misal pada kolom 2 nilai paling tinggi adalah diarah East, maka langkah yang akan diambil adalah kearah east.

**HASIL PEMROSESAN**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |  |  | **O** | **O** |
|  |  |  |  |  |  |  |  |  |  |  |  | **O** | **O** |  |
|  |  |  |  |  |  |  |  |  |  |  | **O** | **O** |  |  |
|  |  |  |  |  |  |  |  |  |  |  | **O** |  |  |  |
|  |  |  |  |  |  |  |  |  | **O** | **O** | **O** |  |  |  |
|  |  |  |  |  |  |  |  |  | **O** |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  | **O** |  |  |  |  |  |
|  |  |  |  |  |  | **O** | **O** | **O** | **O** |  |  |  |  |  |
|  |  |  |  | **O** | **O** | **O** |  |  |  |  |  |  |  |  |
|  |  |  |  | **O** |  |  |  |  |  |  |  |  |  |  |
|  |  | **O** | **O** | **O** |  |  |  |  |  |  |  |  |  |  |
|  | **O** | **O** |  |  |  |  |  |  |  |  |  |  |  |  |
|  | **O** |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **O** | **O** |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **O** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |



Pada hasil diatas, terdapat step step yang dilakukan. Proses perjalanan dapat dilihat pada proses diatas. Reward yang didapatkan hingga sampai ke titik goals adalah 448.