Pattern Recognition Methods
and
Introduction to Machine Learning

Homework 8 - Report
Q Witcher

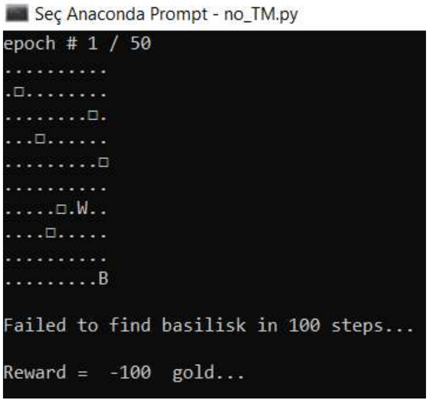
Prepared by: Mehmet Kapson

02.06.2019

To implement the Q learning for this situation, quest map (environment) is created with wanted properties. 10x10 board (100 states) is chosen as quest map. There are impassible mountains which block some ways (states). The Witcher (which is of course Geralt of Rivia) is initially located in 0x0 (state 0). Besides, basilisk is located at the end of the quest map (10x10, state 99). If Geralt finds the basilisk, he gets a big reward (100 gold). Each step gives him -1 gold as a reward; thus, he wants to find the basilisk in a shortest way to slay it. Moreover, if he cannot find the basilisk in 100 steps, he will fail the mission (finish the epoch) and he will start it again (start new epoch). After all of the epochs, q-table is written to an excel file. For every situation, epoch = 50. After each epoch, epsilon is decreased by using this equation, epsilon -= decay*epsilon. Thus, while our character's knowledge about quest map is growing, he takes less random actions. For other words, he less explores but more exploits. In the quest map, W represents the Witcher, □ represents impassible mountains, B represents the basilisk, x represents toxic mist.

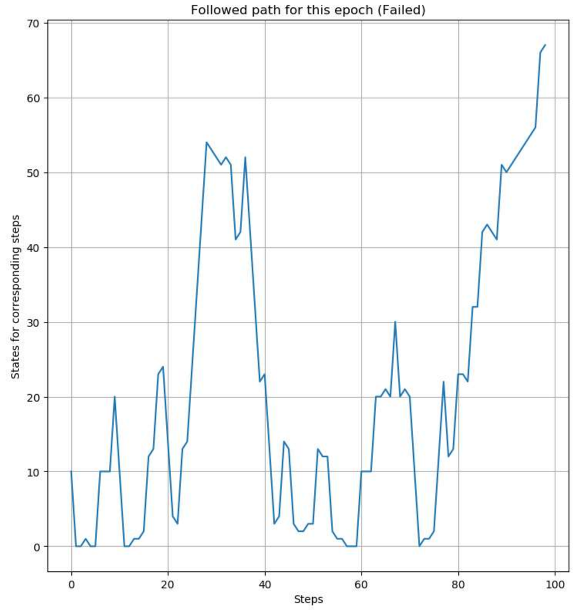
Obtained Results

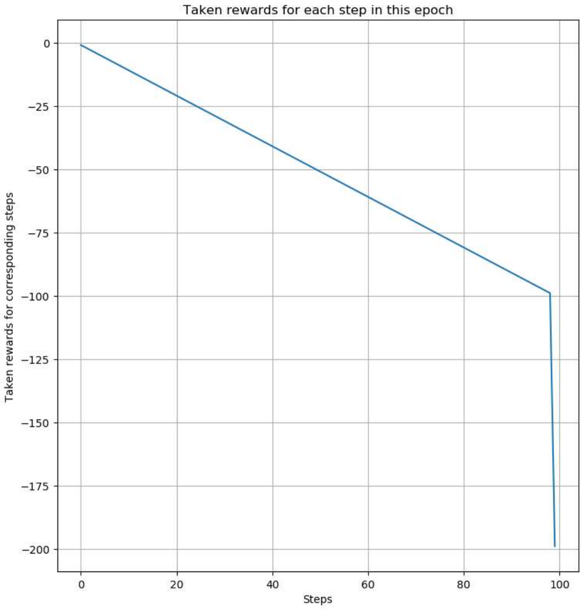
1) There are not any toxic mists in quest map. alpha = 1, gamma = 0.1, epsilon = 0.1, decay = 0.1. He took 100 steps and could not find it. He got -100 gold in his first run.



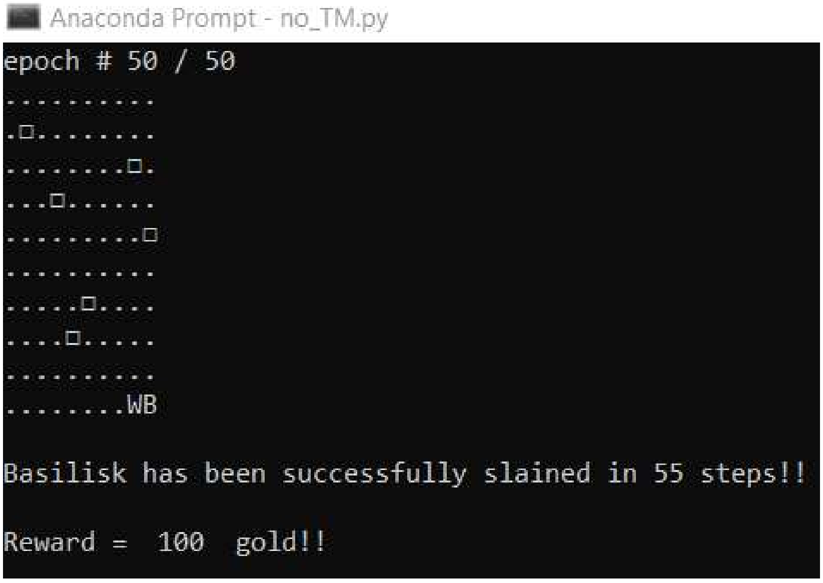
```
epoch # 1 / 50
..........
.□........
........□.
...□......
.........□
..........
.....□.W..
....□.....
..........
.........B

Failed to find basilisk in 100 steps...

Reward =  -100  gold...
```

Followed path and taken reward for this epoch is shown in these graphs.


Followed path for this epoch (Failed)

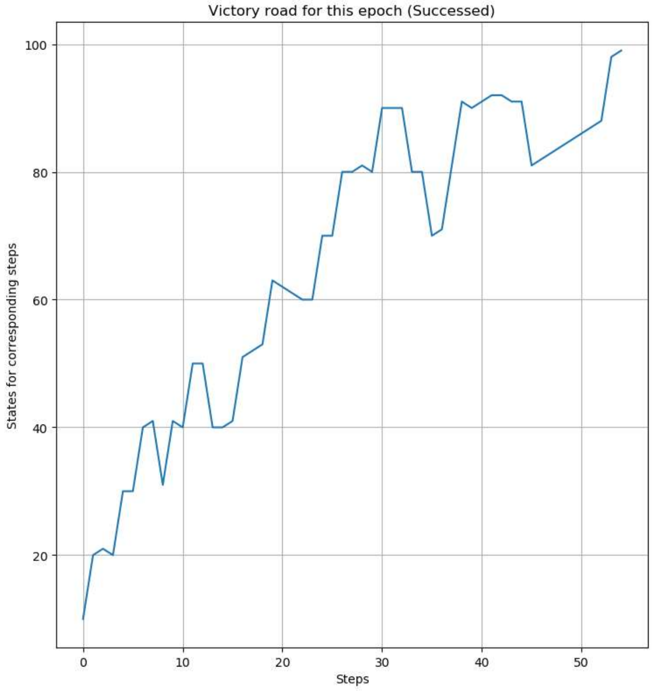

Taken rewards for each step in this epoch

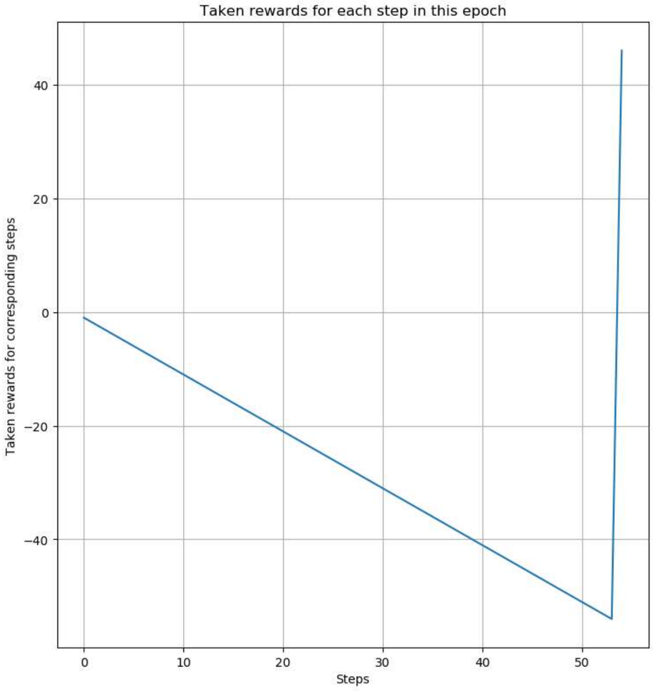
He trained with 50 epochs and these are the results for last (fifth) epoch. He found and slayed the basilisk in 55 steps.



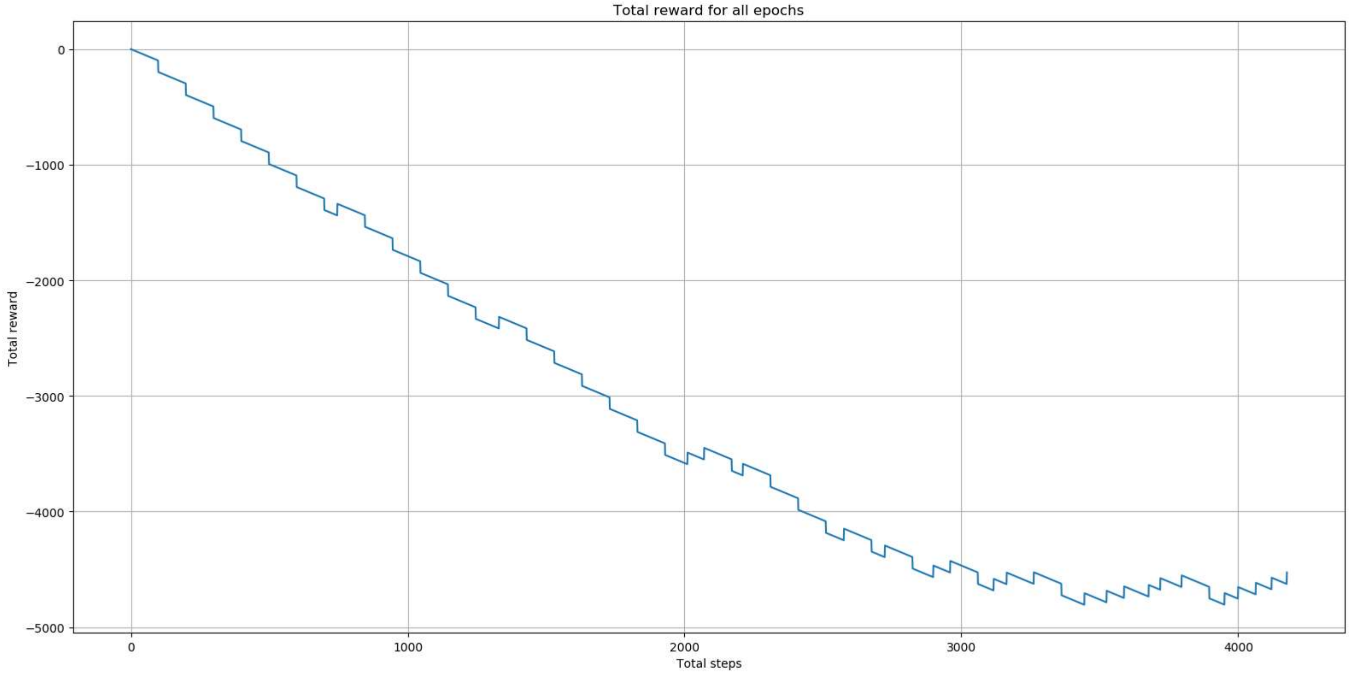
```
epoch # 50 / 50
..........
.□........
........□.
...□......
.........□
..........
.....□....
....□.....
..........
........WB

Basilisk has been successfully slained in 55 steps!!

Reward =  100  gold!!
```

Followed path and taken reward for this epoch is shown below in these graphs.



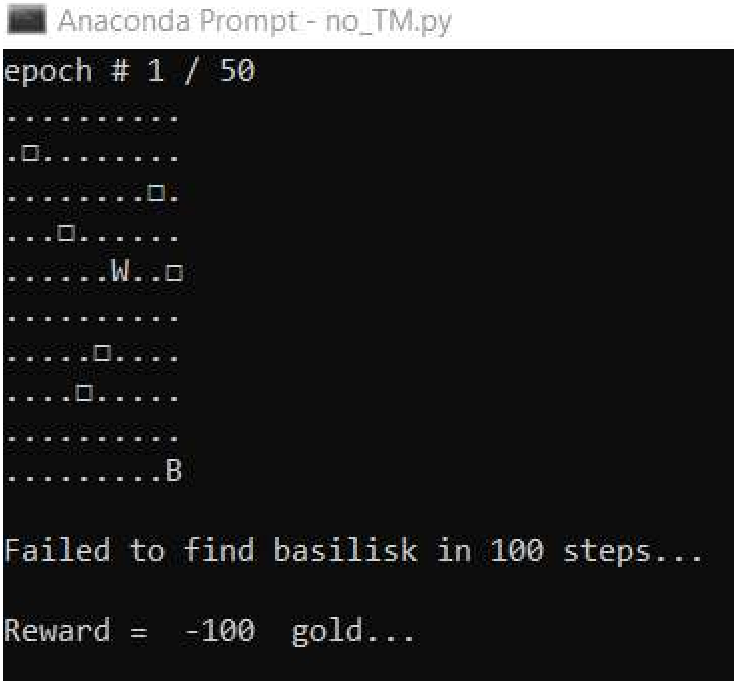


Total reward for all epochs is given as a graph below.



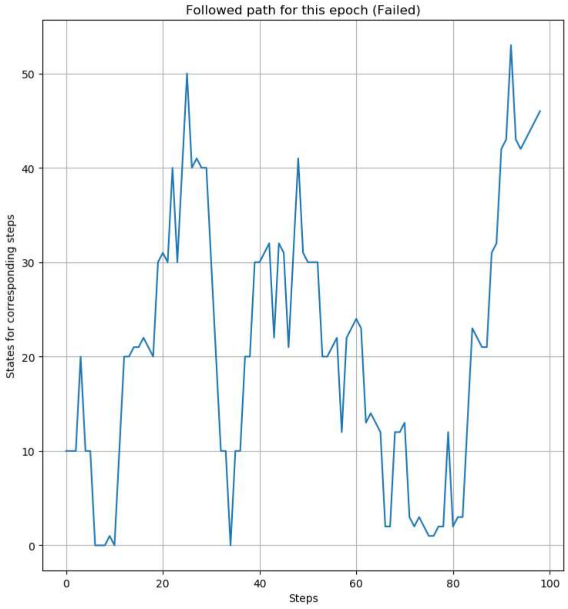
He completed the epochs with -reward. That means he could not find the basilisk for most of the epochs.

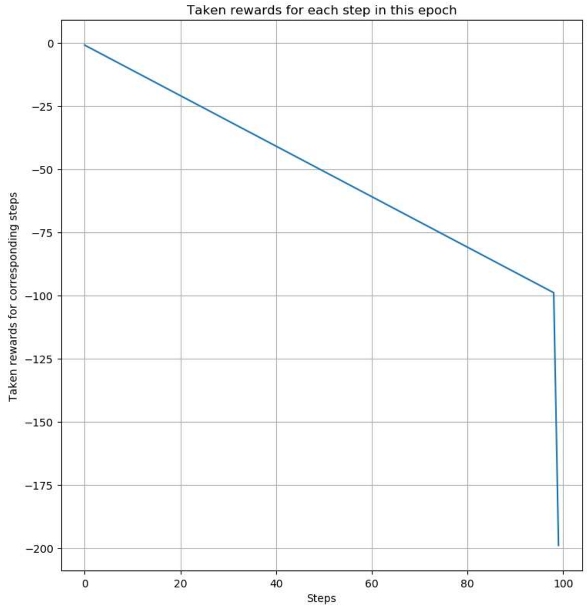
2) There are not any toxic mists in quest map. alpha = 0.8, gamma = 0.1, epsilon = 0.1, decay = 0.1. Just alpha (learning rate) is decreased to 0.8. He took 100 steps and could not find it. He got -100 gold in his first run.



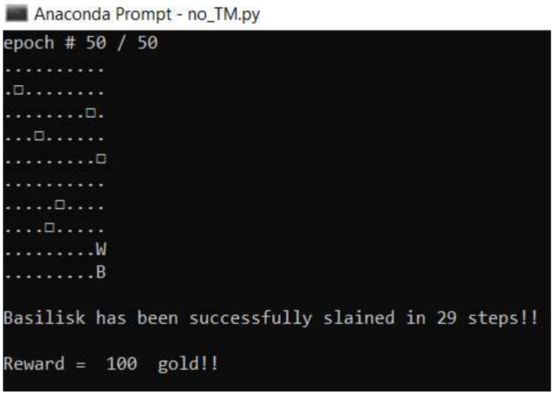
```
epoch # 1 / 50
..........
.□........
........□.
...□......
......W..□
..........
.....□....
....□.....
..........
.........B

Failed to find basilisk in 100 steps...

Reward =  -100  gold...
```

Followed path and taken reward for this epoch is shown below in these graphs.



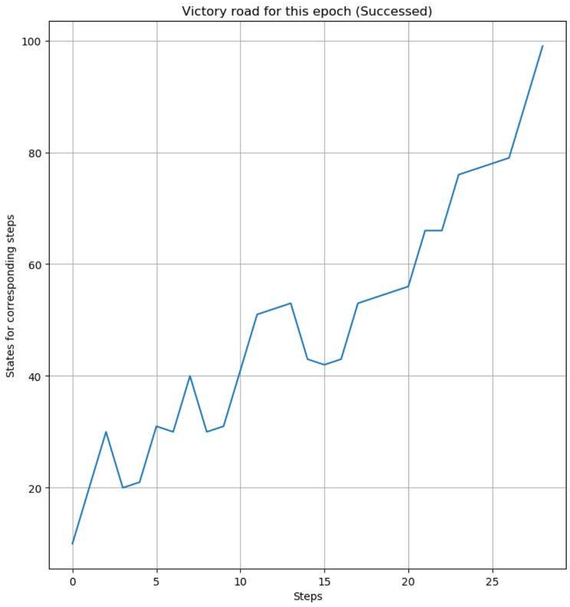


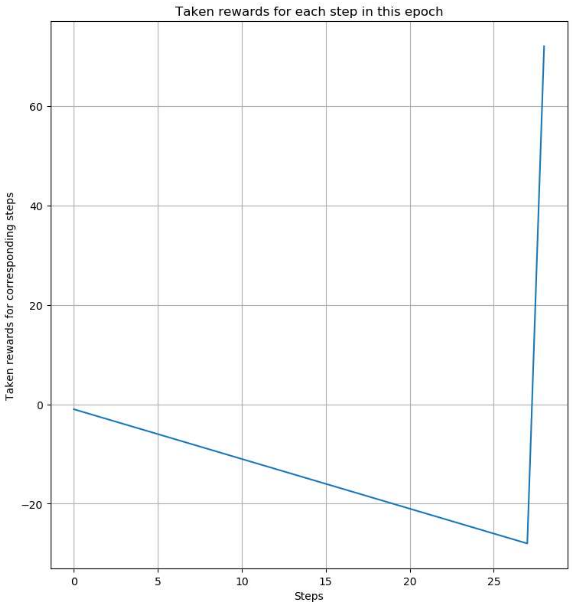
He trained with 50 epochs and these are the results for last (fifth) epoch. He found and slayed the basilisk in 29 steps.



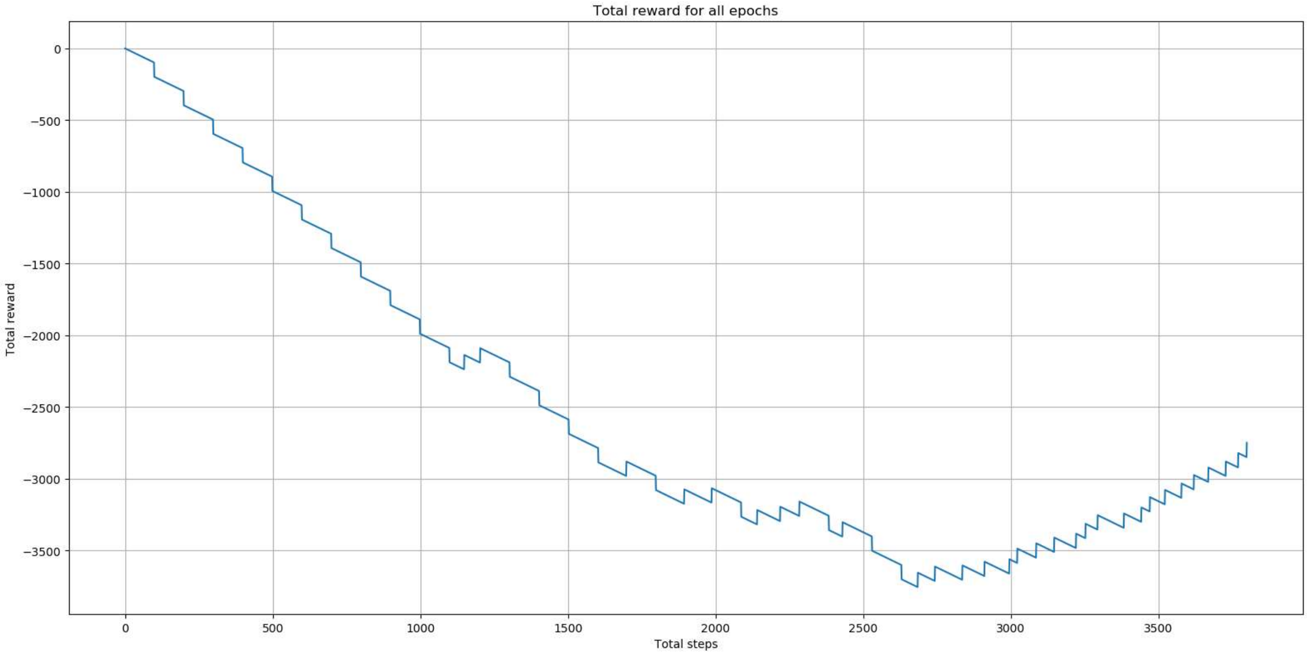
```
epoch # 50 / 50
..........
.□........
........□.
...□......
.........□
..........
.....□....
....□.....
.........W
.........B

Basilisk has been successfully slained in 29 steps!!

Reward =  100  gold!!
```

Followed path and taken reward for this epoch is shown in these graphs.



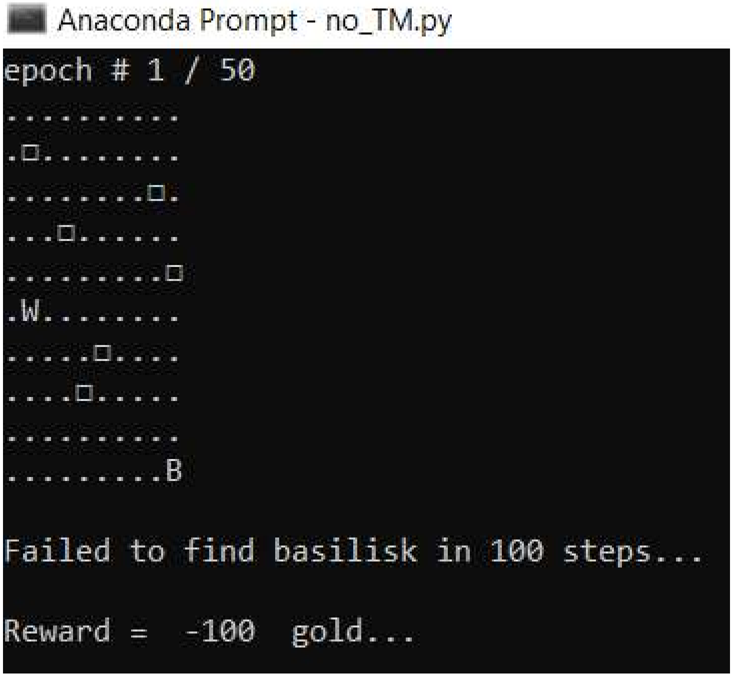


Total reward for all epochs is given as a graph below.



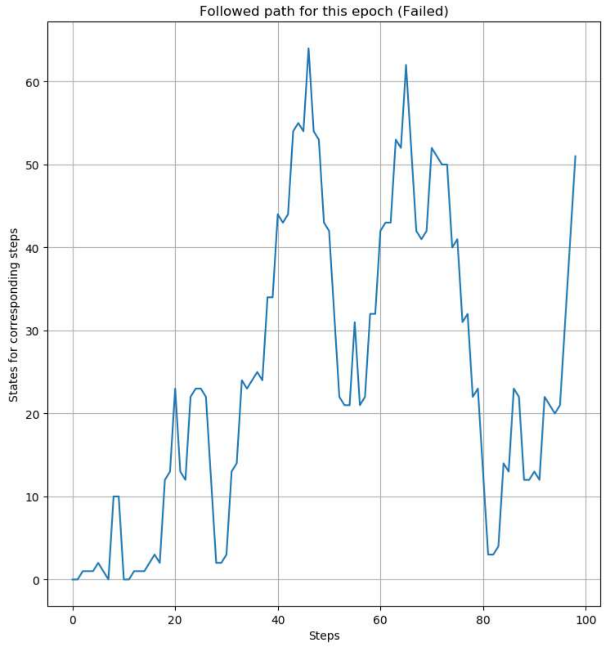
As one can see that, while just alpha is reduced from 1 to 0.8, reward is bigger than previous experiment. Besides, steps that taken is reduced. All of these means that Geralt found the basilisk more often than previous experiment. With alpha = 0.8, our witcher can explores different situations.

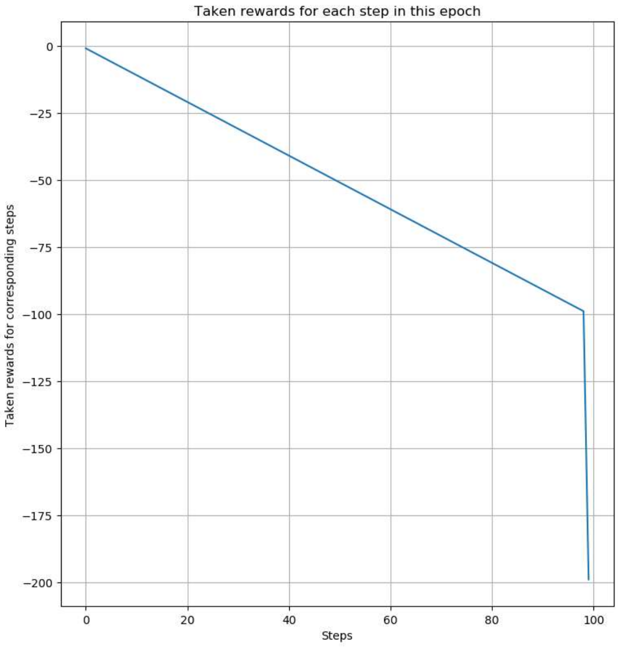
3) Again, there are not any toxic mists in quest map. alpha = 0.8, gamma = 0.05, epsilon = 0.1, decay = 0.1. Just gamma (discount factor) is decreased to 0.05. He took 100 steps and could not find it. He got -100 gold in his first run.



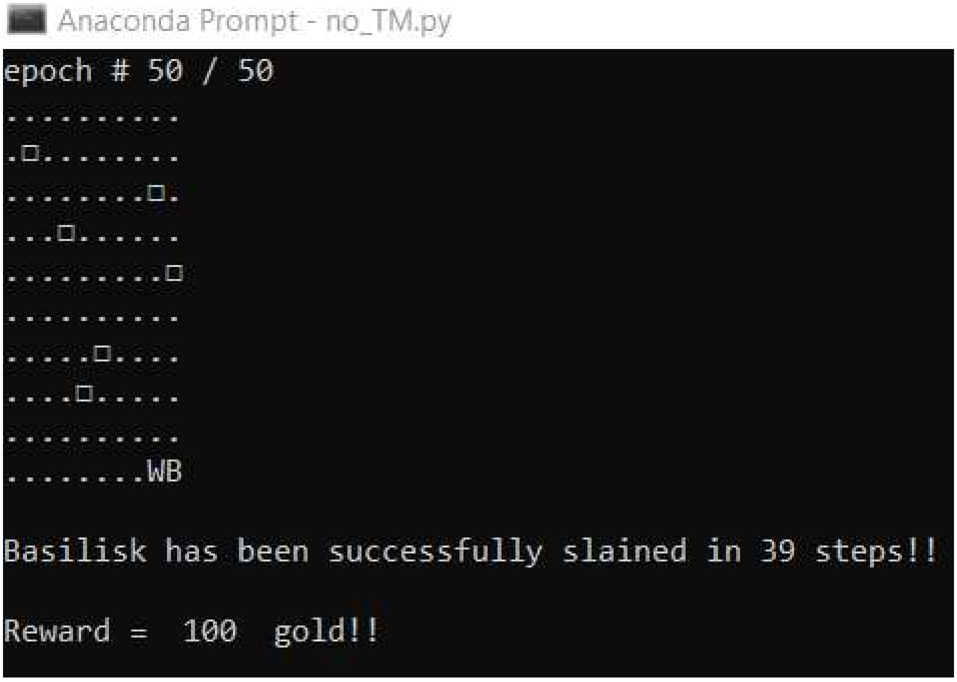
```
epoch # 1 / 50
..........
.□........
........□.
...□......
.........□
.W........
.....□....
....□.....
..........
.........B

Failed to find basilisk in 100 steps...

Reward =  -100  gold...
```

Followed path and taken reward for this epoch is shown in these graphs.



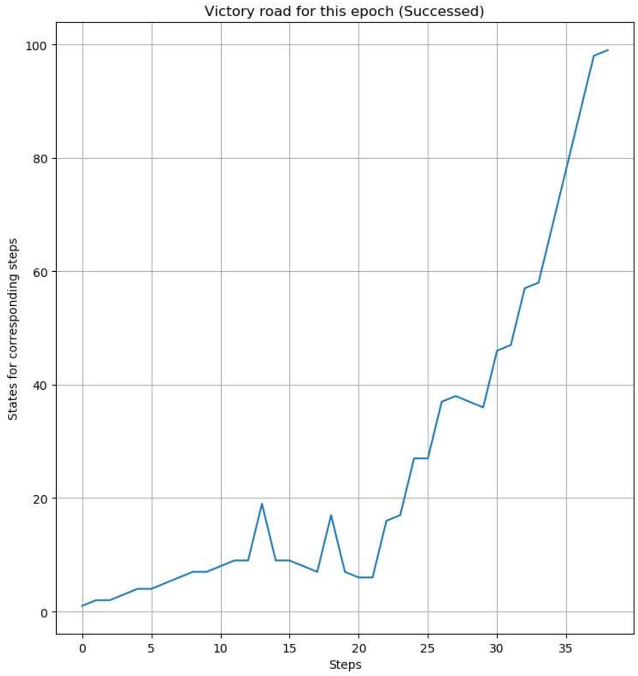


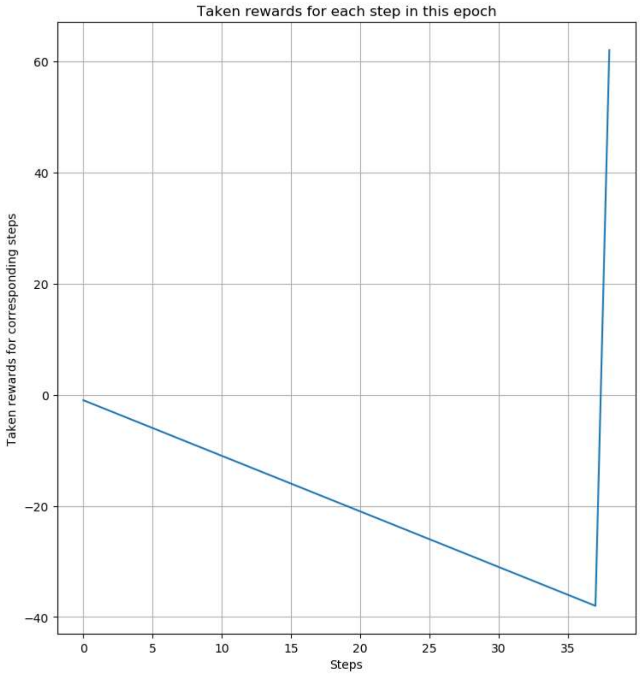
He trained with 50 epochs and these are the results for last (fifth) epoch. He found and slayed the basilisk in 39 steps.



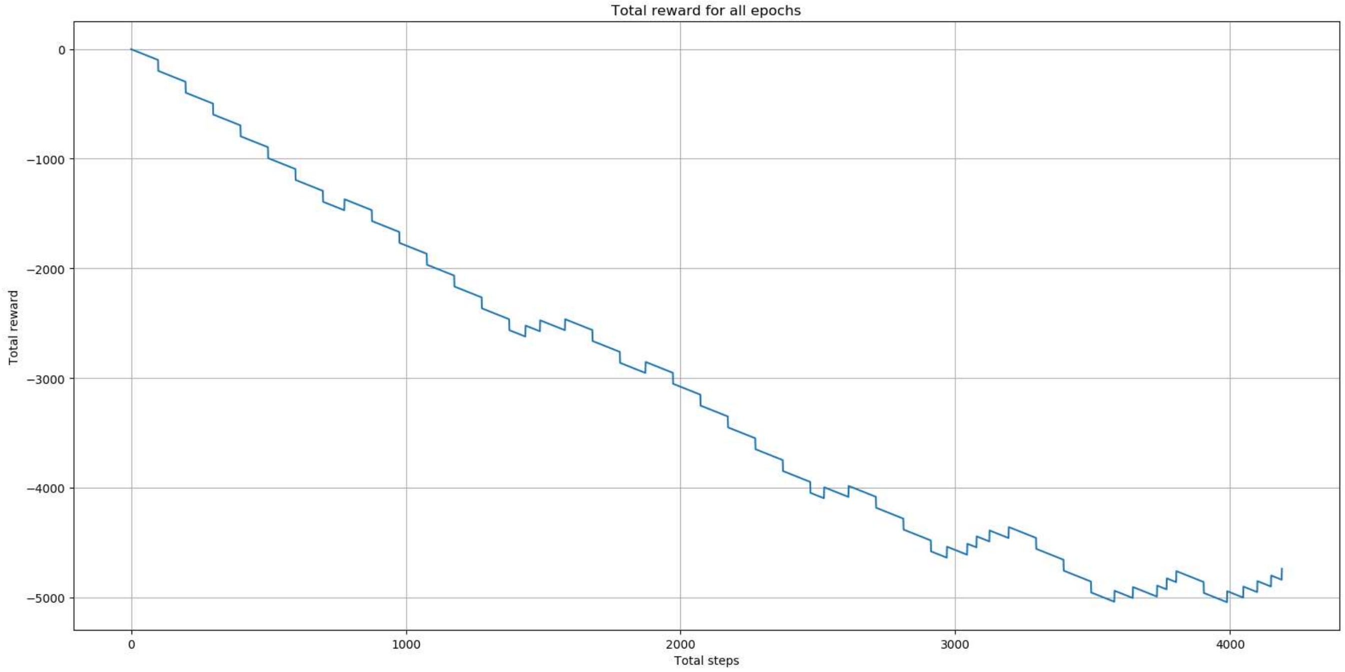
```
epoch # 50 / 50
..........
.□........
........□.
...□......
.........□
..........
.....□....
....□.....
..........
........WB

Basilisk has been successfully slained in 39 steps!!

Reward =  100  gold!!
```

Followed path and taken reward for this epoch is shown in these graphs.



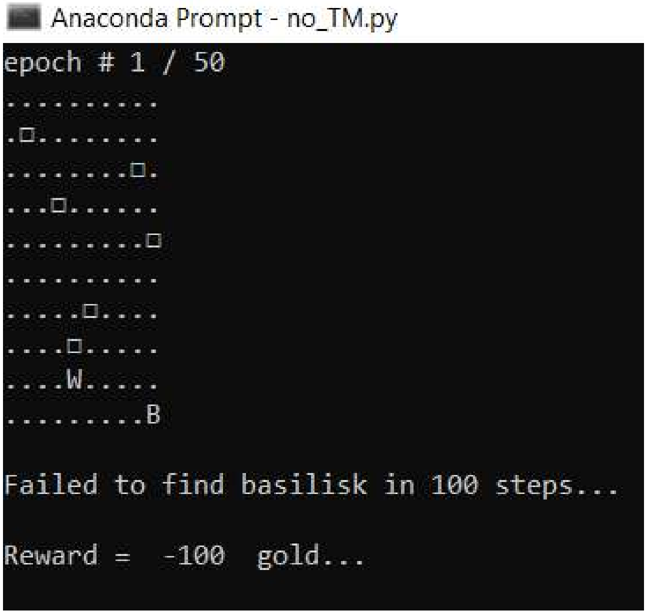


Total reward for all epochs is given as a graph below.



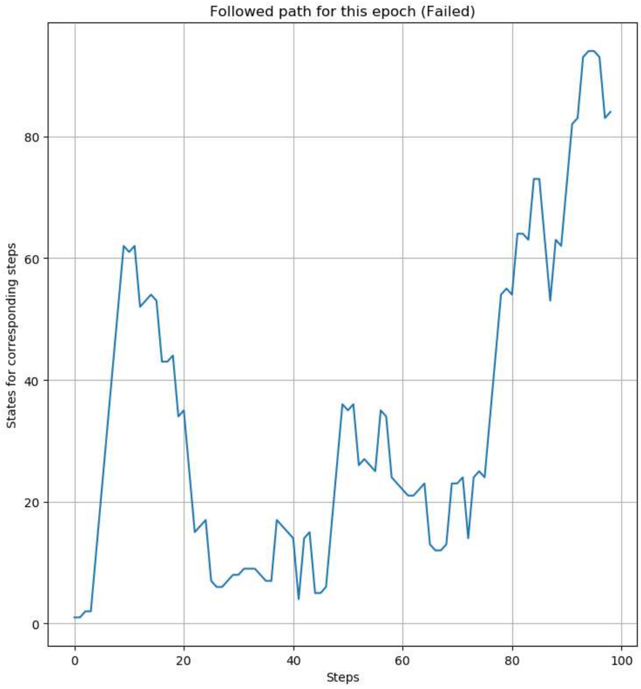
As one can see that, while just gamma is reduced from 0.1 to 0.05, reward is smaller than in 2nd situation. Besides, steps that taken is increased. All of these means that Geralt found the basilisk less often than in 2nd situation. He mostly sought for current rewards, he did not think about long term rewards that much. Moreover, he still got -reward like previous situations.

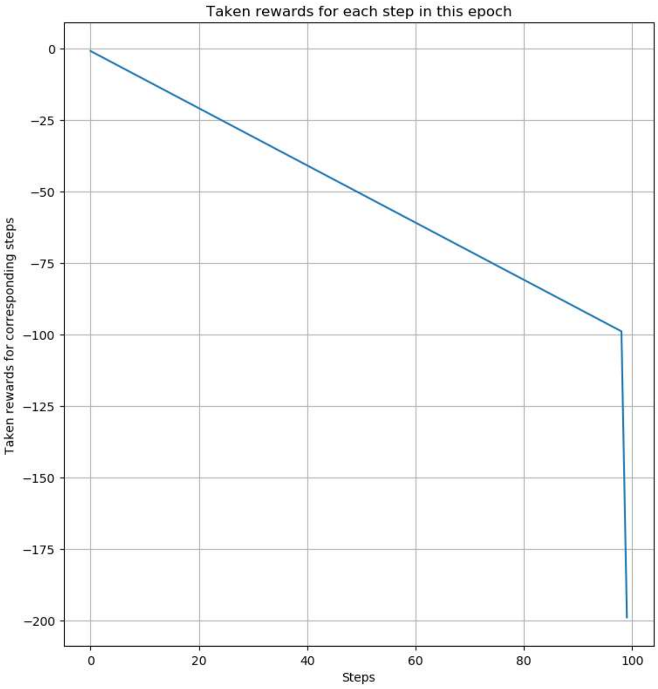
4) Again, there are not any toxic mists in quest map. alpha = 0.8, gamma = 0.99, epsilon = 0.1, decay = 0.1. Just gamma (discount factor) is decreased to 0.05. He took 100 steps and could not find it. He got -100 gold in his first run.



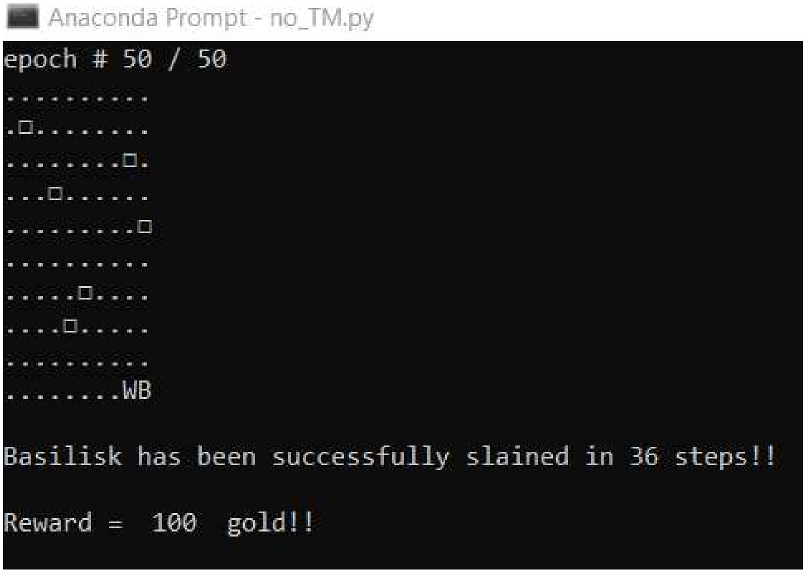
```
epoch # 1 / 50
..........
.□........
........□.
...□......
.........□
..........
.....□....
....□.....
....W.....
.........B

Failed to find basilisk in 100 steps...

Reward =  -100  gold...
```

Followed path and taken reward for this epoch is shown in these graphs.



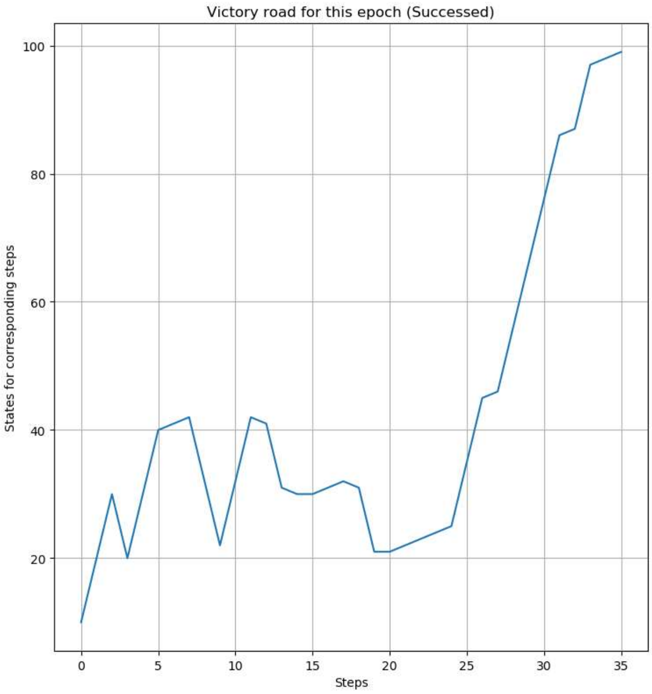


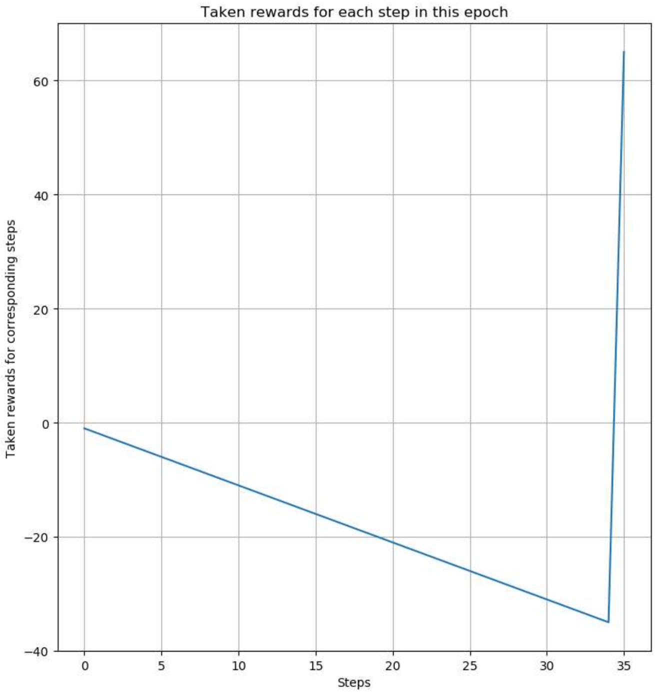
He trained with 50 epochs and these are the results for last (fifth) epoch. He found and slayed the basilisk in 36 steps. Although he finished last epoch with 36 steps, the shortest path which is 18 steps is seen in epoch #48.



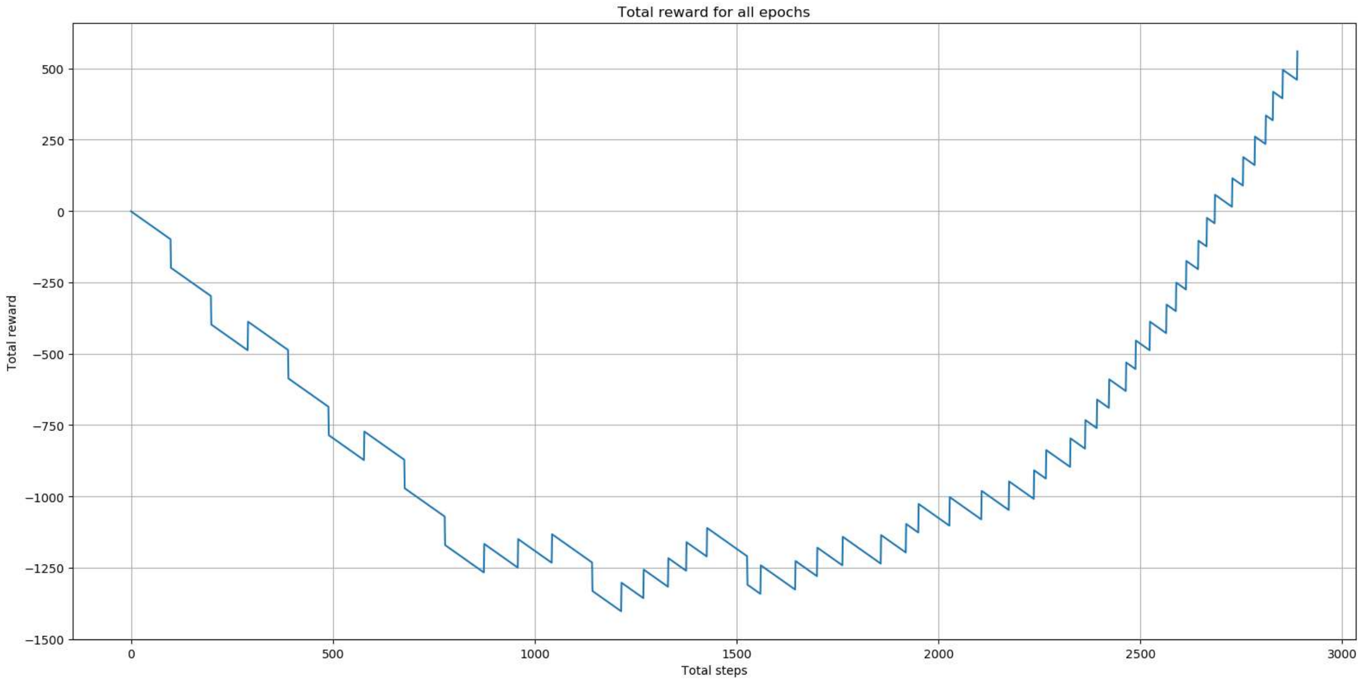
```
epoch # 50 / 50
..........
.□........
........□.
...□......
.........□
..........
.....□....
....□.....
..........
........WB

Basilisk has been successfully slained in 36 steps!!

Reward =  100  gold!!
```

Followed path and taken reward for this epoch is shown in these graphs.



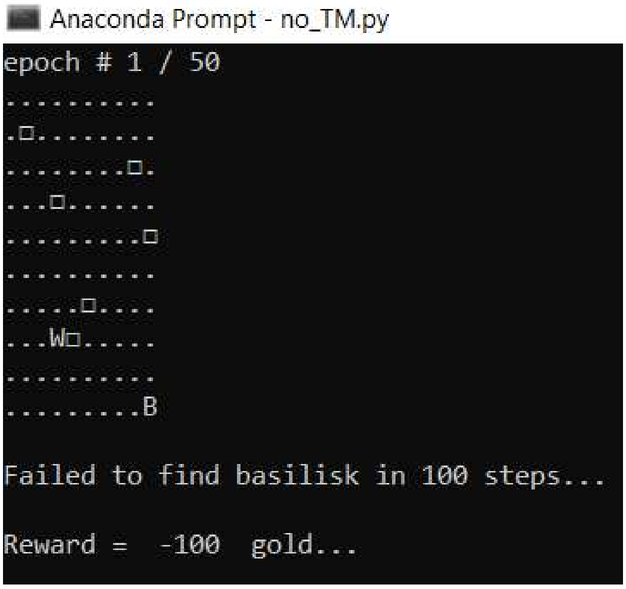


Total reward for all epochs is given as a graph below.



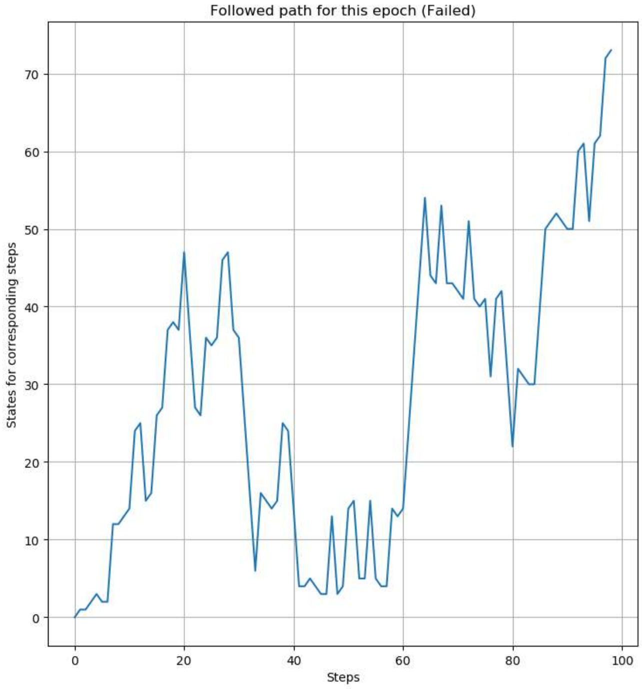
As one can see that, while just gamma is increased from 0.05 to 0.99, reward became much greater than in 3rd situation. Besides, steps that taken is decreased. He mostly sought for long term rewards. Moreover, he finally got around +565 gold as a reward instead of getting negative reward like previous situations. All of these means that Geralt found the basilisk most of the time. This is the best result so far.

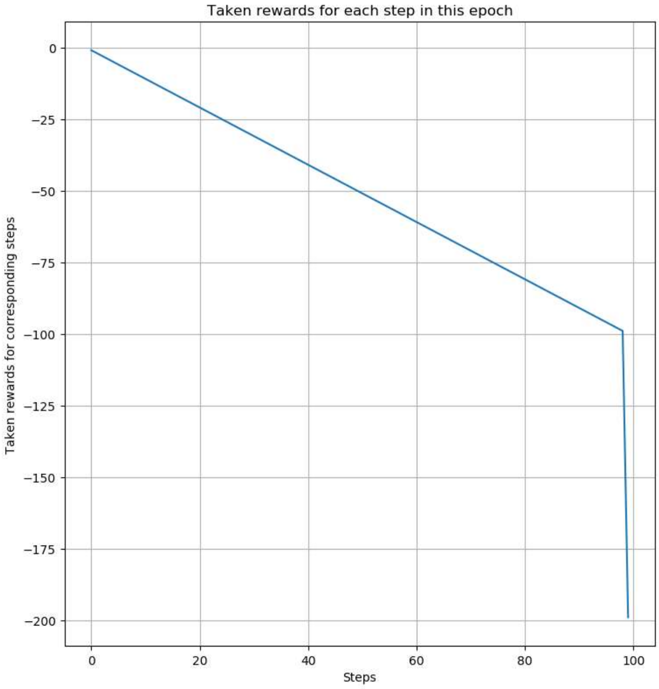
5) Again, for the last time to check epsilon's effect, there are not any toxic mists in quest map. alpha = 0.8, gamma = 0.99, epsilon = 0.3, decay = 0.1. Just epsilon (experiment rate) is increased to 0.3. He took 100 steps and could not find it. He got -100 gold in his first run.



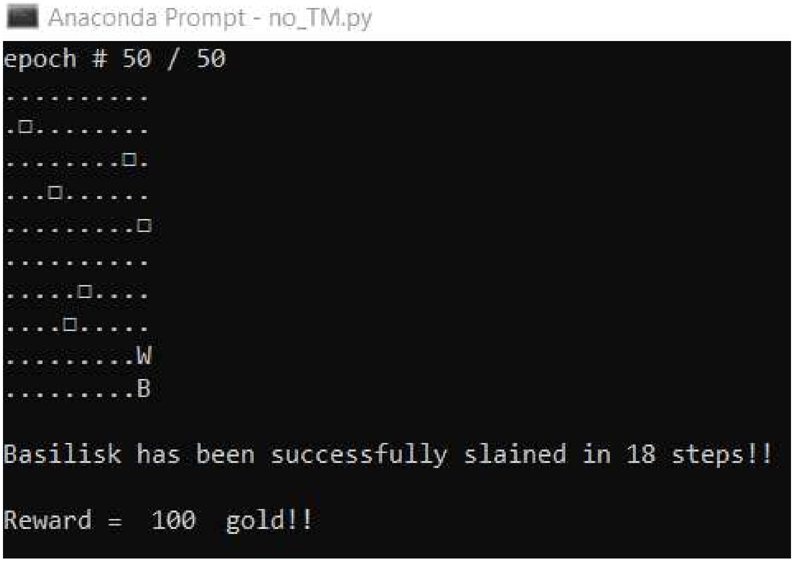
```
epoch # 1 / 50
..........
.□........
........□.
...□......
.........□
..........
.....□....
...W□.....
..........
.........B

Failed to find basilisk in 100 steps...

Reward =  -100  gold...
```

Followed path and taken reward for this epoch is shown in these graphs.



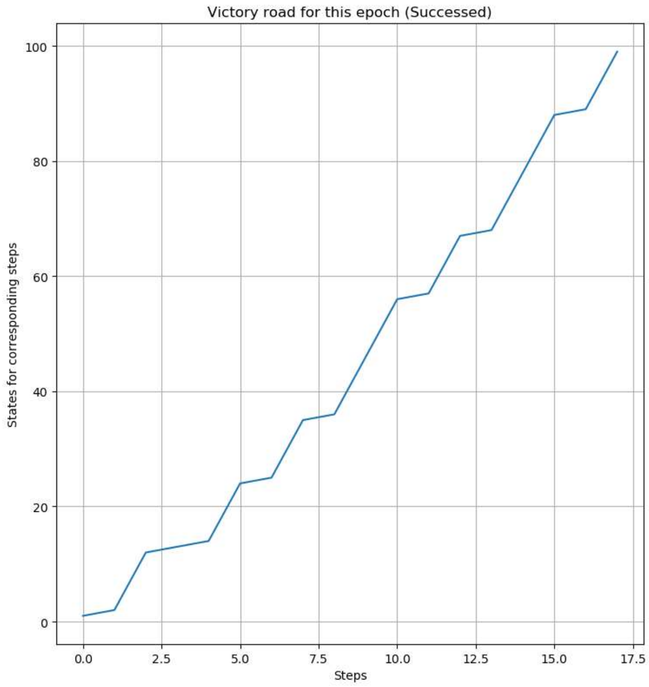


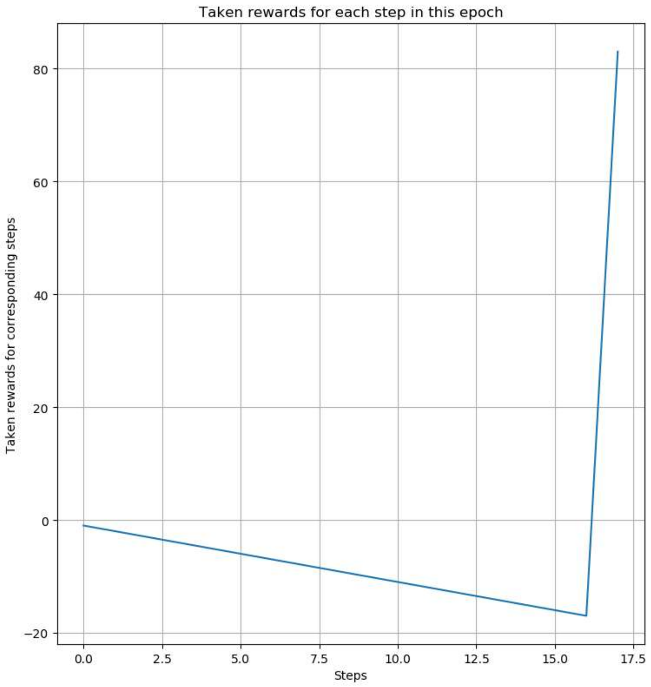
He trained with 50 epochs and these are the results for last (fifth) epoch. He found and slayed the basilisk in 18 steps.



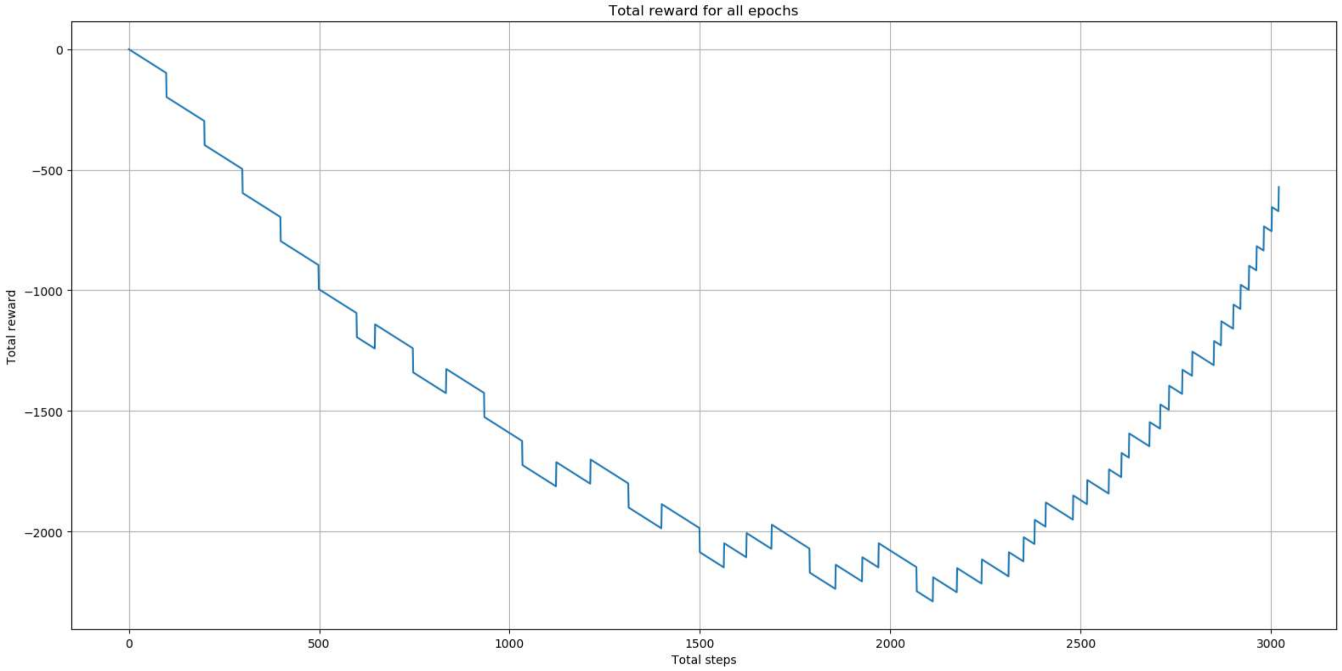
```
epoch # 50 / 50
..........
.□........
........□.
...□......
.........□
..........
.....□....
....□.....
.........W
.........B

Basilisk has been successfully slained in 18 steps!!

Reward =  100  gold!!
```

Followed path and taken reward for this epoch is shown in these graphs.



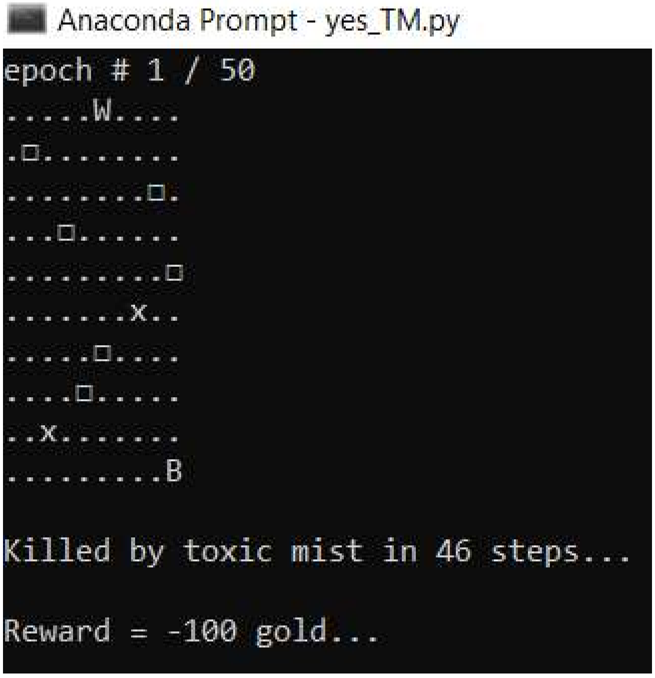


Total reward for all epochs is given as a graph below.



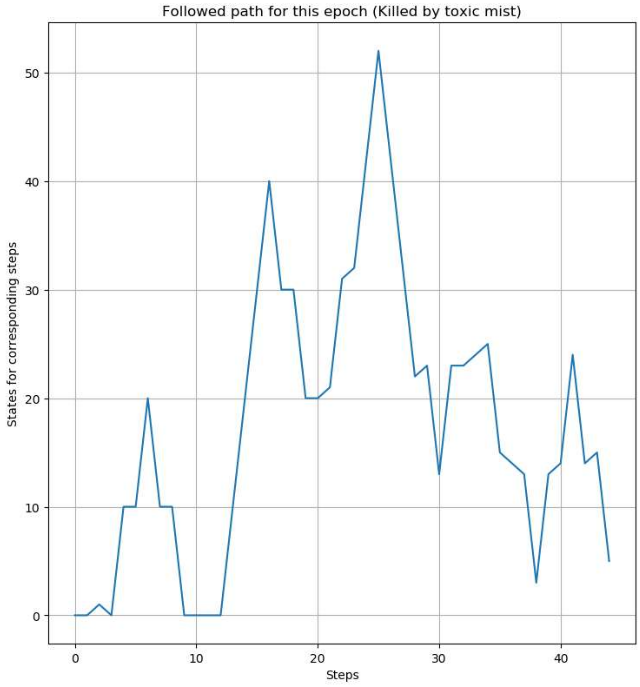
Increasing epsilon from 0.1 to 0.3 affected the results in a bad way. This parameter helps our witcher to decide whether takes a random action or not. Besides, increasing it causes to take more random action choice rather than a safer action. He got negative reward at the end even though he found basilisk in 18 steps for his last epoch.

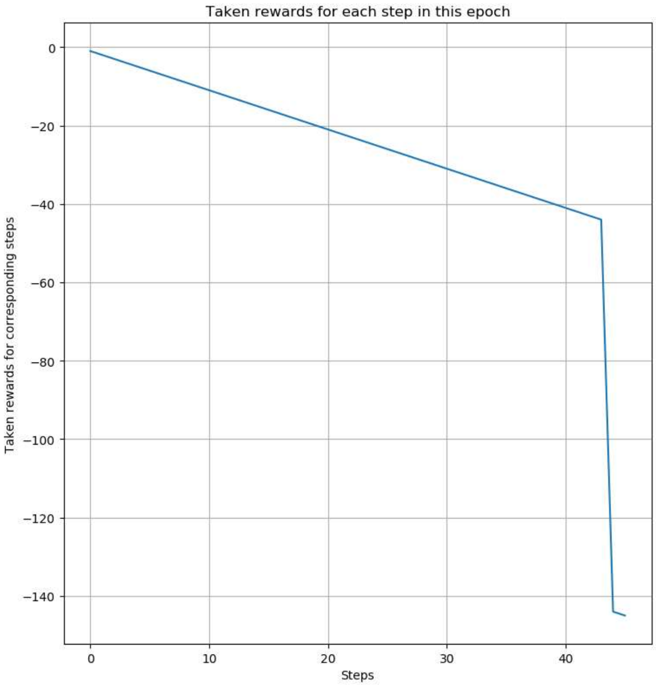
6) For this time, there are toxic mists (represented by ‘x’) in 3 different states in quest map. If he steps any of these toxic mists, he will die. Therefore, new epoch will start from the beginning. Besides, he will get -100 gold as a reward. alpha = 0.8, gamma = 0.99, epsilon = 0.1, decay = 0.1, which gave the best result in 4th situation. He took 46 steps and unfortunately killed by toxic mist. He got -100 gold in his first run.



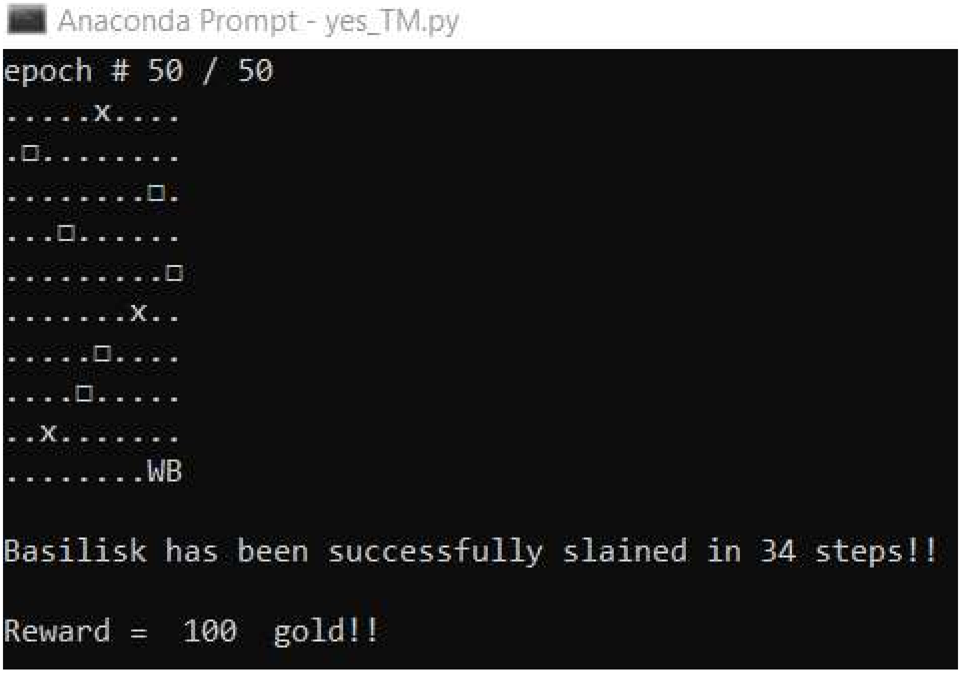
```
epoch # 1 / 50
.....W....
.□........
........□.
...□......
.........□
.......x..
.....□....
....□.....
..x.......
.........B

Killed by toxic mist in 46 steps...

Reward = -100 gold...
```

Followed path and taken reward for this epoch is shown in these graphs.



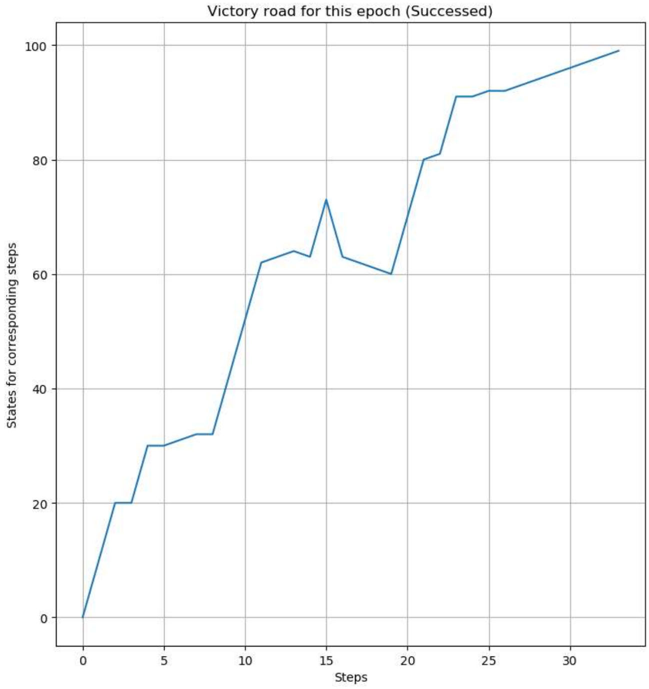


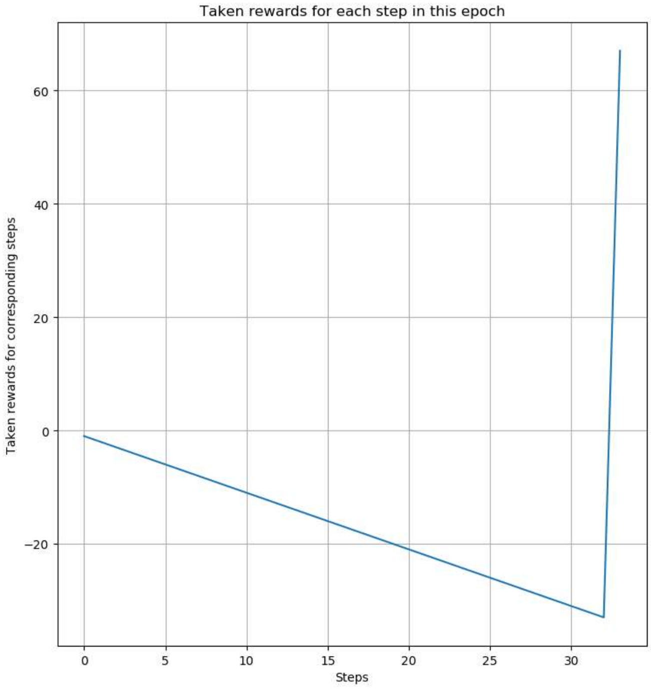
He trained with 50 epochs and these are the results for last (fifth) epoch. He found and slayed the basilisk in 34 steps. In epoch # 49, he found it in 18 steps.



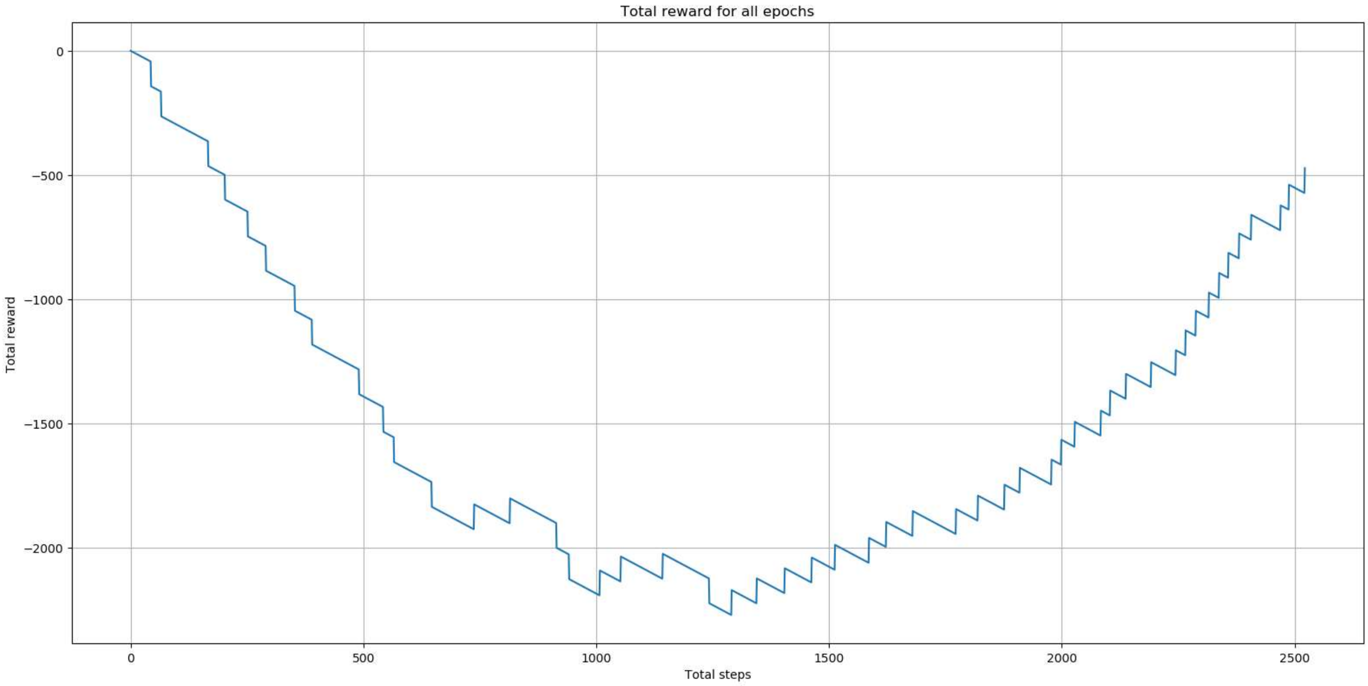
```
epoch # 50 / 50
.....x....
.□........
........□.
...□......
.........□
.......x..
.....□....
....□.....
..x.......
........WB

Basilisk has been successfully slained in 34 steps!!

Reward =  100  gold!!
```

Followed path and taken reward for this epoch is shown in these graphs.



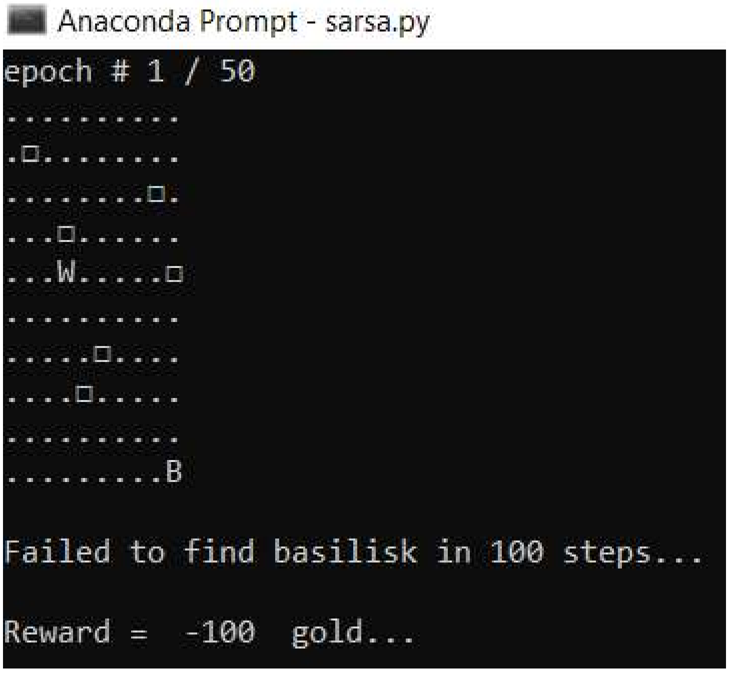


Total reward for all epochs is given as a graph below.



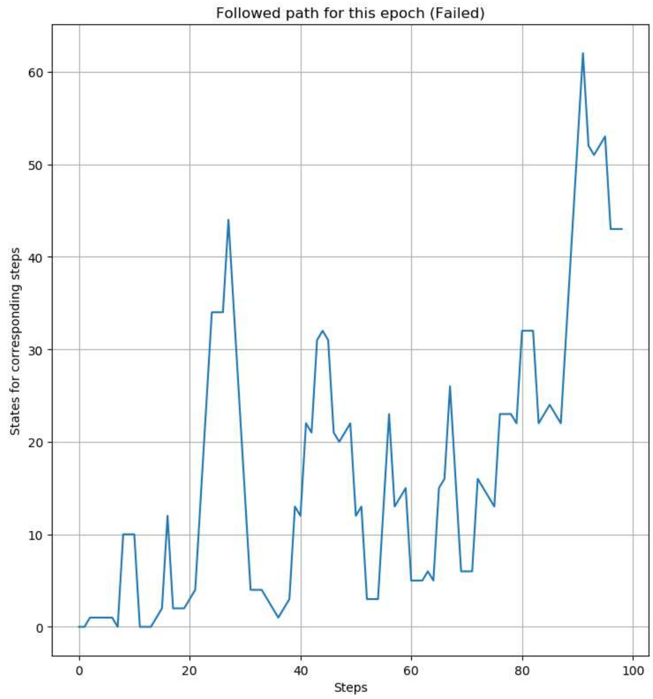
In 4th situation with the same parameters but without the toxic mists, total reward was positive and around 565 gold. For this situation, total reward is negative. While he was learning that toxic mists kill him, he got lots of negative reward. Once he learned this information by trying and trying again, he did not step for toxic mists and he finished the epochs successfully with positive reward.

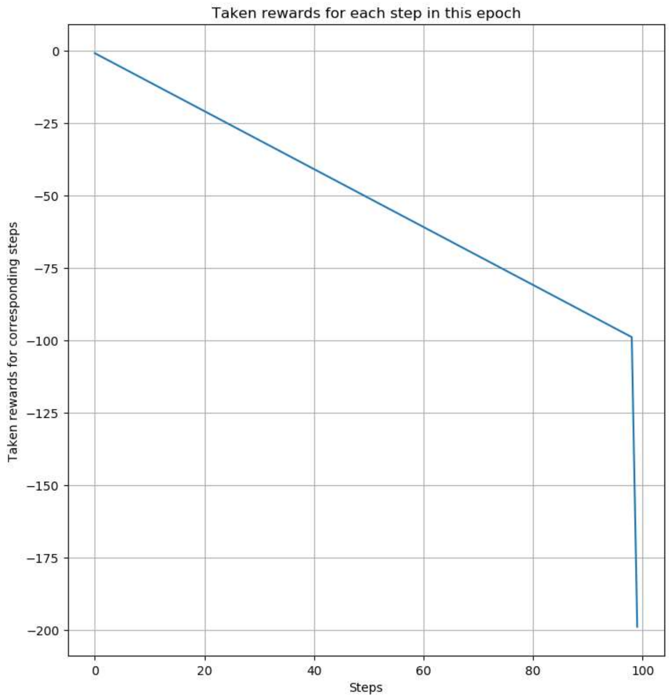
7) For this time, there are not any toxic mists but SARSA algorithm is applied instead of Q-Learning. alpha = 0.8, gamma = 0.99, epsilon = 0.1, decay = 0.1, which gave the best result in 4th situation in Q-Learning. He took 100 steps. He got -100 gold in his first run.



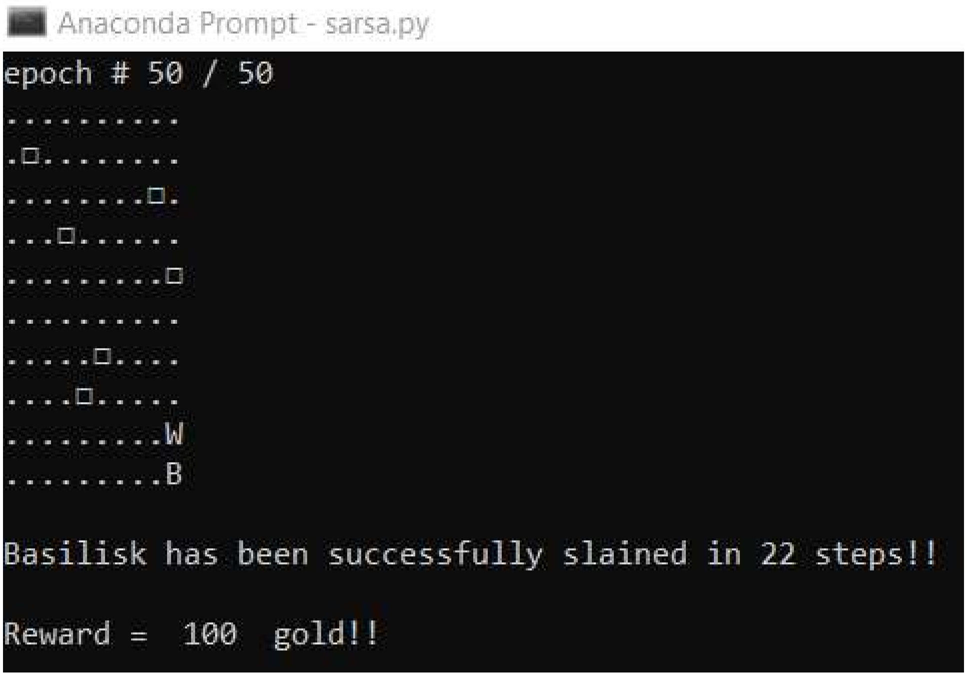
```
epoch # 1 / 50
..........
.□........
........□.
...□......
...W.....□
..........
.....□....
....□.....
..........
.........B

Failed to find basilisk in 100 steps...

Reward =  -100  gold...
```

Followed path and taken reward for this epoch is shown in these graphs.



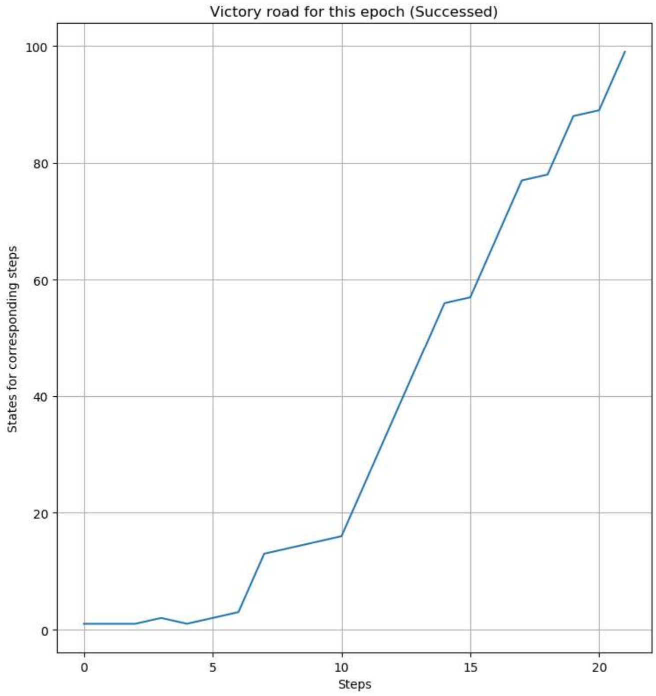


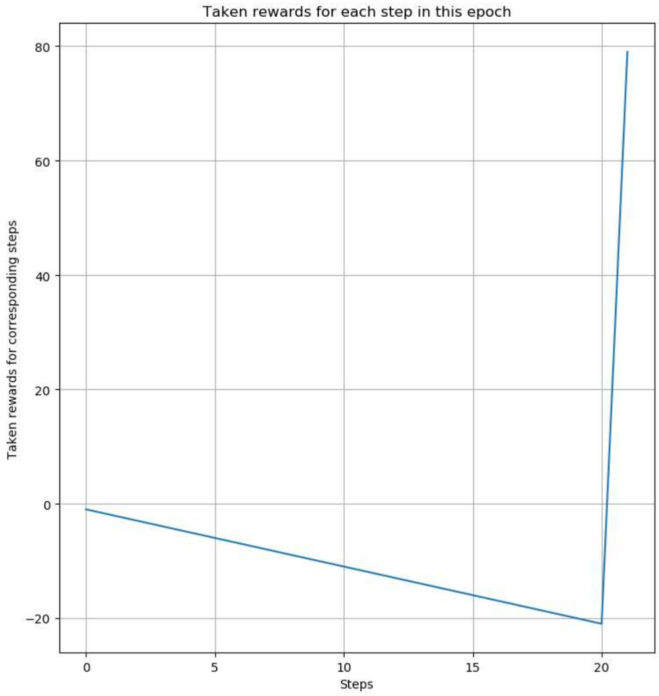
He trained with 50 epochs and these are the results for last (fifth) epoch. He found and slayed the basilisk in 22 steps.



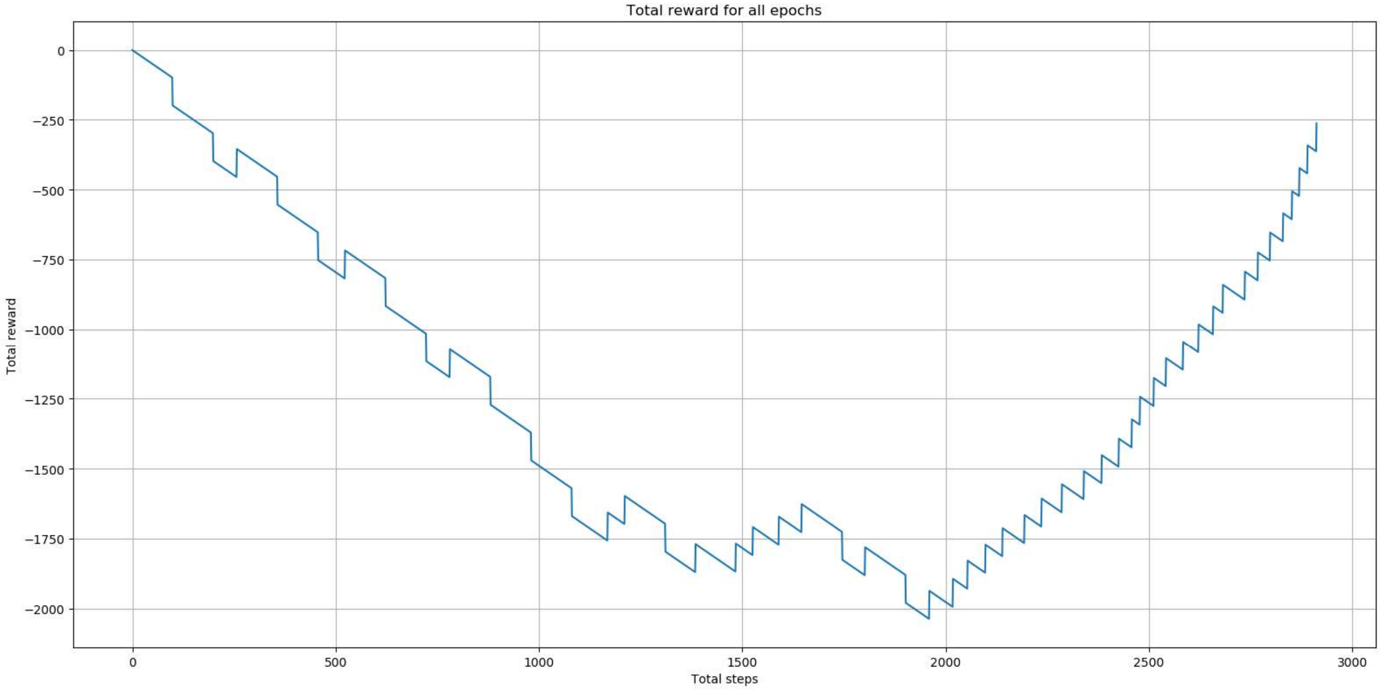
```
epoch # 50 / 50
..........
.□........
........□.
...□......
.........□
..........
.....□....
....□.....
.........W
.........B

Basilisk has been successfully slained in 22 steps!!

Reward =  100  gold!!
```

Followed path and taken reward for this epoch is shown in these graphs.



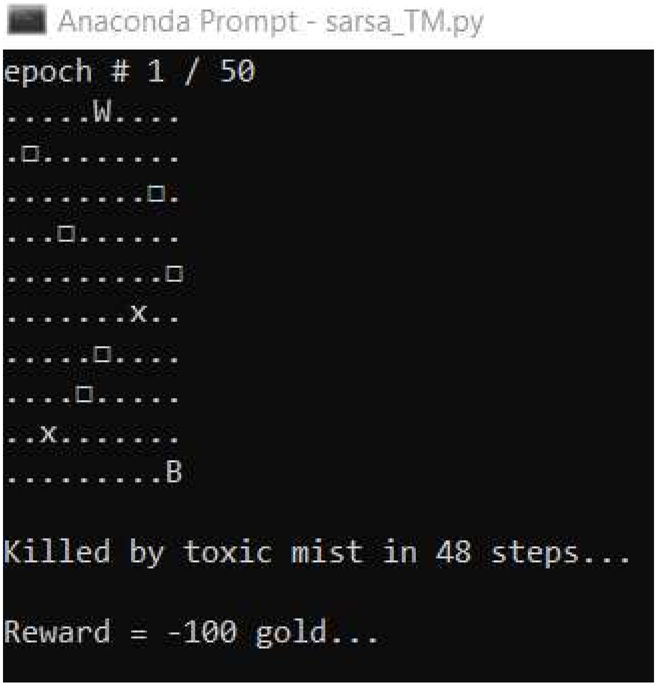


Total reward for all epochs is given as a graph below.



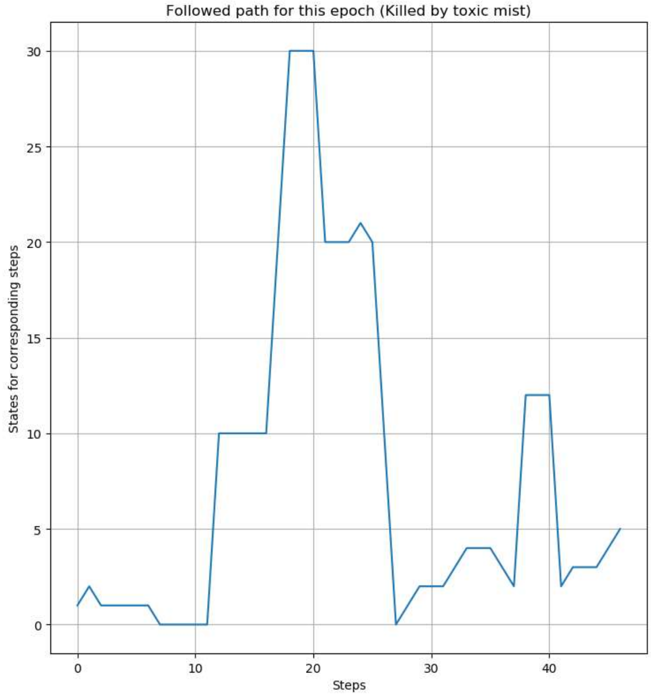
As one can see that total reward is negative while in q learning it was positive. In SARSA, our witcher has on-policy to take his next action.

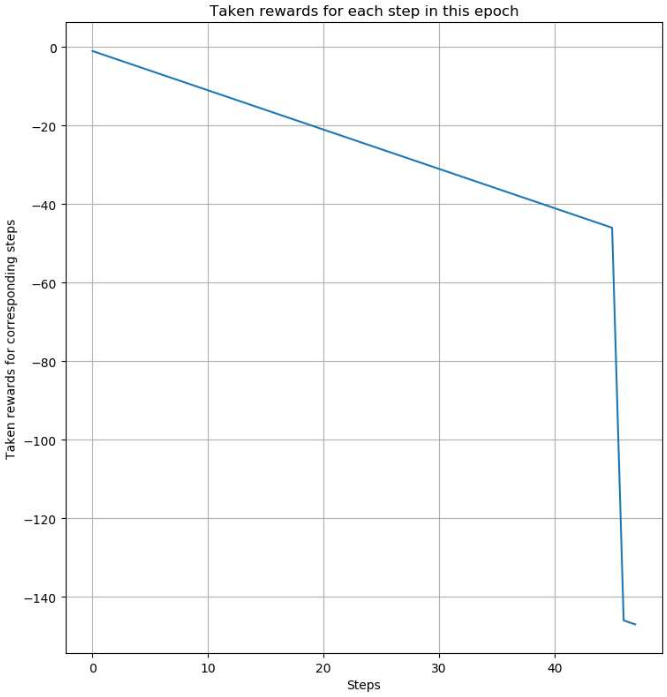
8) For this time, toxic mists are added in same states again and SARSA algorithm is applied instead of Q-Learning again. alpha = 0.8, gamma = 0.99, epsilon = 0.1, decay = 0.1, which gave the best result in 4th situation in Q-Learning. He took 78 steps and unfortunately killed by toxic mist. He got -100 gold in his first run.



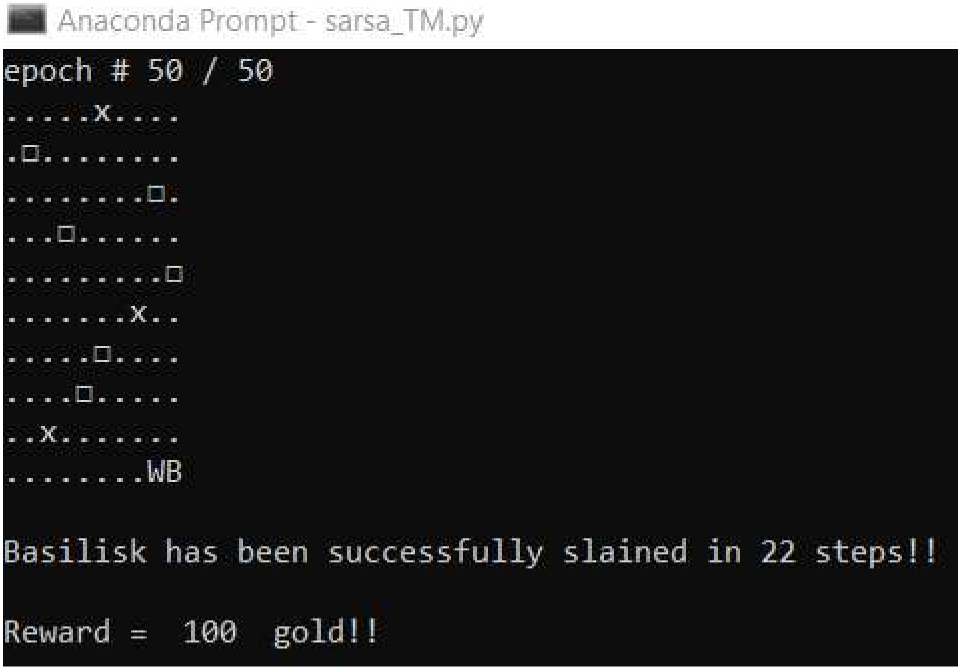
```
epoch # 1 / 50
.....W....
.□........
........□.
...□......
.........□
.......x..
.....□....
....□.....
..x.......
.........B

Killed by toxic mist in 48 steps...

Reward = -100 gold...
```

Followed path and taken reward for this epoch is shown in these graphs.



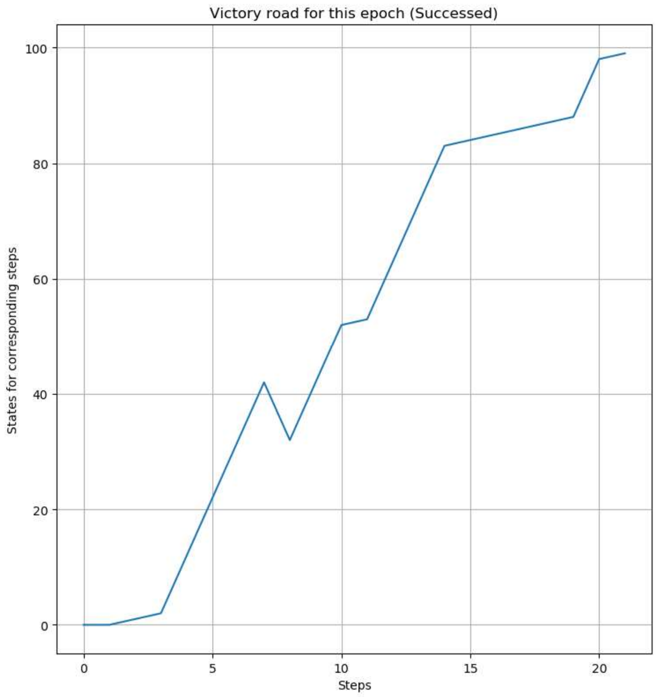


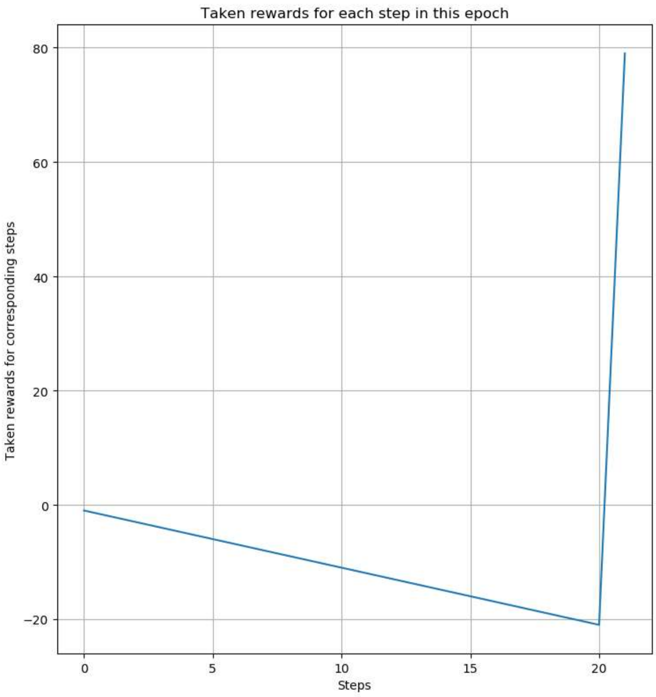
He trained with 50 epochs and these are the results for last (fifth) epoch. He found and slayed the basilisk in 22 steps.



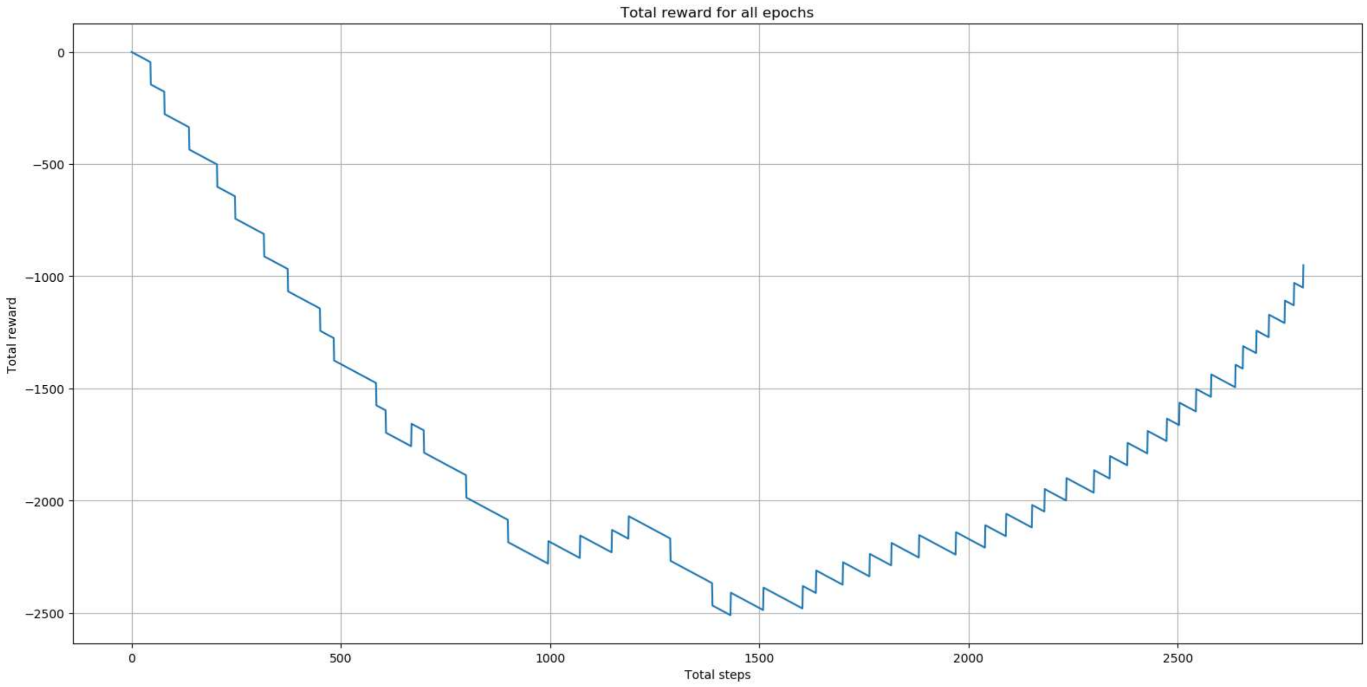
```
epoch # 50 / 50
.....x....
.□........
........□.
...□......
.........□
.......x..
.....□....
....□.....
..x.......
........WB

Basilisk has been successfully slained in 22 steps!!

Reward =  100  gold!!
```

Followed path and taken reward for this epoch is shown in these graphs.





Total reward for all epochs is given as a graph below.



The results are worse than Q-Learning. He got around -500 gold as a total reward with Q-Learning for the same parameters. With SARSA, he got around -1000 gold as a total reward. Q-Learning worked better.

Conclusion

To sum up, Q-Learning and SARSA algorithms are applied for both map with or without toxic mists. Different alpha, gamma and epsilon parameters are used to see the effects of each one. Comments are added after each situation and result. Q-Learning appears to be better than SARSA in these experiments and maps.

P.S: Q-Tables are added as different excel files for each situation.