AI Assignment-II
Team-50

Nikhil Rayaprolu-201501090
Vegulla Sri Kavya-201501062

For us X=50
delta=X/20=50/20=2.5
gamma=1
reward=-X/20=-50/20=-2.5

Matrices after every iteration of value iteration algorithm:-

Iteration 1:

0	0	50	0
-2.5	-2.5	37.5	-2.5
-2.5	-50	0	-2.5
-2.5	-2.5	-2.5	-2.5

Iteration 2:

0	0	50	0
-5	22.25	37	27
-5	-50	0	-5
-5	-5	-5	-5

Iteration 3:

0	0	50	0
14.3	24.325	42.425	29.3
-7.5	-50	0	18.1
-7.5	-7.5	-7.5	-7.5

Iteration 4:

0	0	50	0
17.64	28.8725	42.8625	36.18
3.19	-50	0	24.56
-10	-10	-10	10.48

Iteration 5:

0	0	50	0
22.681	29.6773	44.0053	37.864
6.931	-50	0	31.356
-1.948	-12.5	3.884	17.196

Iteration 6:

0	0	50	0
24.203	30.6719	44.2541	39.6262
11.3379	-50	0	34.0624
1.6	-5.6428	12.0336	24.6928

Iteration 7:

0	0	50	0
25.5916	30.9705	44.5298	40.2722
12.9962	-50	0	36.0134
6.16604	1.5626	19.661	28.4226

Iteration 8:

0	0	50	0
26.1352	31.2209	44.6243	40.7524
14.2729	-50	0	36.9204
8.66982	8.38503	24.1702	31.1191

Iteration 9:

0	0	50	0
26.5175	31.3215	44.6973	40.9667
14.8354	-50	0	37.486
10.6238	12.6747	27.2293	32.5653

Iteration 10:

0	0	50	0
26.6925	31.39	44.7288	41.1031
15.1976	-50	0	37.7706
11.6982	15.5509	28.9981	33.4683

Iteration 11:

0	0	50	0
26.801	31.4221	44.7493	41.1704
15.3738	-50	0	37.9366
12.6303	17.2536	30.0742	33.9631

Results for delta =0

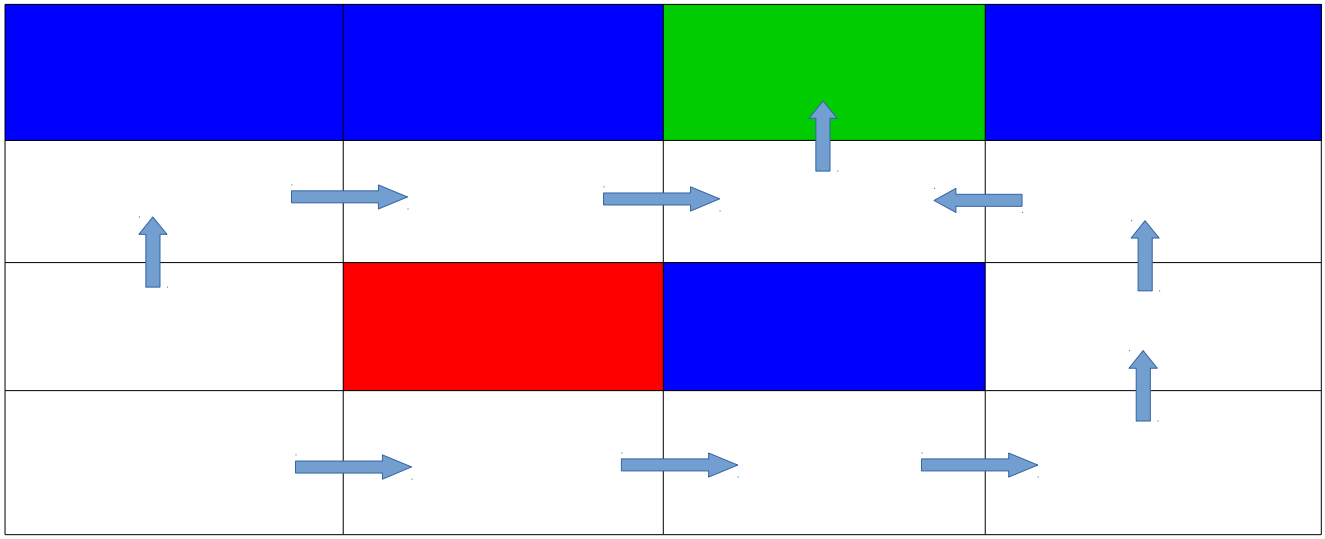
0	0	50	0
26.9233	31.464	44.772	41.2564
15.5985	-50	0	38.1313
16.4258	19.6564	31.4896	34.6153

Expected Reward:-

Final Expected Reward=*12.6303(**From value iteration algorithm****)***

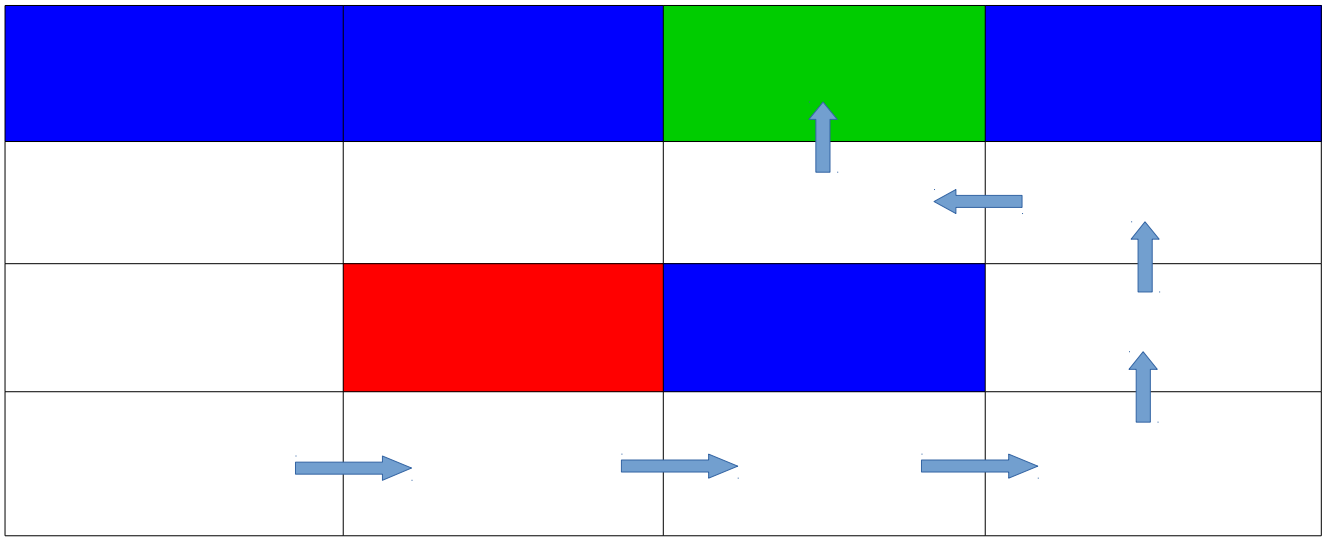
=16.4258 *(if delta=0)*

***Optimal Policy Table*:-**



***Optimal Path*:-**

From start state(3,0) to terminal state(0,2) is
(3,0) -> (3,1) -> (3,2) ->(3,3) -> (2,3) -> (1,3) -> (1,2) -> (0,2)



Linear Programing

***Values of X*:**

State,Action pair	Value of X
3,5	0.864702571
5,1	0
5,2	0
5,3	0
5,4	0.121765601
6,1	0
6,2	0
6,3	0
6,4	0.228333669
7,1	1.080878214
7,2	0
7,3	0
7,4	0
8,1	0
8,2	0
8,3	1.122764098
8,4	0
9,1	0.136986301
9,2	0
9,3	0
9,4	0

10,5	0.135297429
12,1	1.127999833
12,2	0
12,3	0
12,4	0
13,1	0
13,2	0
13,3	0
13,4	1.111111111
14,1	0
14,2	0
14,3	0
14,4	0.987654321
15,1	0
15,2	0
15,3	0
15,4	1.111111111
16,1	0.987654321
16,2	0
16,3	0
16,4	0

Expected Reward=16.42961063

Description of why the rewards match/don’t match:

*We try to maximise the utility/reward in both methods of solving MDP, so they would both end up achieving the same result if we try make them as accurate as possible. In VI, the reward in the start state is the utility of selecting the best paths possible to terminal states. In LP, the paths we get match the ones in VI so they will also consider similar probabilities. Now the reward in LP is the summation of reward*X for each state,action pair. This will correspond to the values we get in VI if we assume a small delta, since a large delta would not allow enough iterations so that our VI spreads out enough and apporximates the utilities of different states enough times. So if we use a delta not near 0, the vaues in VI and LP might not match as in our case , but on using delta=0 the rewards match in both of them.*