Q Learning

The gamma discount was set to 0.9 the default. The learning rate alpha was set to 0.2. The epsilon was also set to 0.2. The number of training episodes was set to 30000, this was because the policy was expected to converge by then. The Junit test took a total of 3.079s to run, aggressive took 0.921s, defensive took 1.379s and random took 0.779s. The table below shows the results when q-learning agent was running against the other agents.

Opponent	Wins	Losses	Draws
Aggressive	50	0	0
Defensive	27	0	23
Random	50	0	