I am developing project related to reinforcement learning. I need your help for successful completion.  
  
I need to answer following question based on python implementation.so I have to present this task, with meaningful presentation having executable code and visulisation. On top of that I want analysis, which I want to present to prof.

Description-

I have 4X3 matrix grid of stochastic environment. My agent position is most bottom left, and goal position is most top right, my hell state is 3rd raw 2nd column. So agent need to reach goal, without fall in hell state. Agent have 8 actions having directions of N, W, S, E, NW,SW, SE, NE.  
  
methods-  
so above problem could be solve by following methods

1. Policy based iteration
2. Value based iteration
3. generalisation ( policy + value) based

Considerations-

* my agent should run above all methods. And have all statistical data like, no of iteration, and respective values( optimum value) etc
* howmany iterations required for convergence?
* which method is better? Why? Which method is faster? Why?
* environment grid need to be flexible for scaling…for example from 4x3 to 20x20.
* please provide..proper data file folder structure
* Code should be professional object oriented programming, must be error free
* it should be easy to read and explainable language with commenting and autodoc string.
* Code should be numpy based matrix array, and should be flexible variable type, must not hard coded.