Project 1: Home insurance prediction model

Description of problem: The basic idea of this project is to develop a regression model to predict the cost of home insurance for a given house, based on the built year, home price, home location and etc.. In US, every home owner should purpose home insurance for their houses. This model will have a broad interest for the insurance companies / home owner. For home owner, this model will have provide them an idea how much the insurance will cost and evaluate whether the quote from the insurance company is reasonable. For the insurance company, the model might help them to provide a better estimate of the insurance quote.

Concerns: I am not sure if there is enough public home insurance data available. If not, this project can be changed to auto insurance prediction model if there is enough data available.

Project 2: Using the satellite image data to detect oil spill

Description of problem: Deepwater Horizon oil spill in 2010 is the largest marine oil spill in US. 4.9 million barrels of oil were leaked into the Gulf of Mexico and 11 people was killed. It is considered as one of the largest environmental disasters in the world. Correctly and effectively characterizing the spill area is crucial in fighting the spill. The idea of this project to characterize oil spill based on the satellite image. We will utilize the image data during the Deepwater horizon oil spill to develop and train the model. The model can be utilized to automatically detect the future oil spills based on the satellite image.

Project 3: Heart Disease Prediction

Description of problem: The dataset is available in UCI machining learning repository. I am also interested in this problem by using the available datasets to predict the possibilities of having a heart disease. It will help us to be aware of the potential heart disease in advance and thus can be treated as early as possible.

Data: http://archive.ics.uci.edu/ml/datasets/Heart+Disease