**Problem Description**

**For the take home coding exam, you will use the attached patent data set. The variable description is provided in the attached file. Your task is to predict the number of forward citations a patent will get. So your target variable is : fwd.**

**Tasks:**

* Explore your data set. Use visualization if necessary.
* Split data into training and test data set.
* Create a pipeline for data preprocessing steps.
* Use the pipeline to transform train and test data sets
* Apply following Regression models (If your model has a scaling parameter(s) use Grid Search to find the best scaling parameter)
  + KNN regressor
  + Linear regression
  + Linear regression with SGD
  + Ridge
  + Lasso
  + ElasticNet
  + Polynomial regression
  + Polynomial regression with regularization
  + SVM both simple and with kernels (try rbf, poly, and sigmoid kernel),
  + Decision Tree regression.
* The final model should also be chosen based on Cross-Validation Score
* After selecting the best model, report the test and train error based on the chosen best model.

**Deliverables:**

* Use "download as" in the "file" menu to convert your ipython file to a .pdf file
* Submit both .ipynb, and .pdf files to the eLearning

Please do this assignment individually. If we find that student had copied, both the students will be given 0 for this assignment and will be reported to University authorities for further action.