## Persistency 13M Payment Behaviour Prediction

## Problem Description:

Prepare a Machine Learning Model to predict the Persistency 13M Payment Behaviour at the New Business stage.

## Objective:

* Using Machine Learning techniques, provide scores for each policy at the New Business stage the likelihood to pay the 13M premium.
* Identify the segments where maximum non payers are captured

### Dataset:

* “Training” & “Test” Dataset with the raw input attributes and the 13M actual paid/not paid flag.
* “Out of Time” Datasets would be provided with just the raw input attributes.

### Expected Steps:

* Conduct appropriate Data Treatments for e.g. Missing Value Imputation, Outlier treatment etc.
* Conduct required Feature Engineering for e.g. Binning, Ratio, Interaction, Polynomial etc.
* Use any machine learning algorithm or combination of machine learning algorithms you deem fit.
* Prepare your model on the Train Data and you can evaluate the generalization capability of your model by using K-Fold Cross Validation, Leave One Out Cross Validation or any other validation technique that you see appropriate.
* Score the Test and Out of Time Data and share it back to us along with the scored Train Data for evaluation. Also share all the Model Codes and Documentation.

### Evaluation Criteria:

* Maximum non payers captured with minimum population.
* Model should hold across Out of time period
* Mode should be explainable i.e. importance and effect of variables finally selected in the model should be clear
* Characteristics of a good model (try to beat these min criteria, not mandatory)
  + Non Payment Rate in the Bottom Most Probability to Pay Decile should be at least <= 15%.
  + % Non Payers Capture in the Bottom Most Probability to Pay Decile should be at least >= 40%.

### Rules:

* Use Python programming language for the task and feel free to use any framework (Jupyter Notebook / PyCharm / Anaconda Spyder etc.)
* No restriction on any library / package being used.

**The code will be evaluated against an anti-plagiarism tool for code authenticity. Hence, please do not copy paste the exact code from any source directly**