**Credit Risk Prediction**

One popular use of machine learning is the prediction of credit risks in issuing housing loans. This was a major concern during the Lehman crisis[[1]](#footnote-1), when the default of several subprime mortgages causes the eventual collapse of collateral loan. The bank has collected past data on customers’ default behaviour and their characteristics in the data set HMEQ – home equity.

As a newly hired machine learning engineer, you have been tasked to formulate a policy outline (<1 page) to mitigate this risk to the bank. In your outline, you must:

1. Develop a predictive model. You have the choice to select any model in mind. Some options are logistic regression, decision trees, random forests, ensemble models etc.
2. Report your results in a confusion matrix. The bank is most concerned about recall since any default would be detrimental to the bank. As such, you may want to tune your model in this manner.
3. Determine which are the key risk factors that decide whether a customer default or not.
4. Use the model to predict which of the individuals in the hmeq\_pred.csv would probably ‘default’.

The dataset for the project is hmeq.csv. It contains the following columns. Note that some data is missing for certain rows.

**Data Set HMEQ**  
The data set HMEQ reports characteristics and delinquency information for 5,960 home equity loans. A home equity loan is a loan where the obligor uses the equity of his or her home as the underlying collateral. The data set has the following characteristics:  
◾ BAD: 1 = applicant defaulted on loan or seriously delinquent; 0 = applicant paid loan  
◾ LOAN: Amount of the loan request  
◾ MORTDUE: Amount due on existing mortgage  
◾ VALUE: Value of current property  
◾ REASON: DebtCon = debt consolidation; HomeImp = home improvement  
◾ JOB: Occupational categories  
◾ YOJ: Years at present job  
◾ DEROG: Number of major derogatory reports  
◾ DELINQ: Number of delinquent credit lines  
◾ CLAGE: Age of oldest credit line in months  
◾ NINQ: Number of recent credit inquiries  
◾ CLNO: Number of credit lines  
◾ DEBTINC: Debt-to-income ratio

1. On Sept. 15, 2008, [Lehman Brothers](https://www.investopedia.com/terms/l/lehman-brothers.asp) filed for bankruptcy.1﻿ It was a sombre reminder that nothing is forever—even in the richness of the financial and investment worlds.Lehman's demise also made it the largest victim of the U.S. subprime mortgage-induced [financial crisis](https://www.investopedia.com/terms/f/financial-crisis.asp) that swept through global financial markets in 2008. Lehman's collapse was a seminal event that greatly intensified the 2008 crisis and contributed to the erosion of close to $10 trillion in [market capitalization](https://www.investopedia.com/terms/m/marketcapitalization.asp) from global [equit](https://www.investopedia.com/terms/e/equitymarket.asp)y. [↑](#footnote-ref-1)