**Machine Learning - Assignment 1 - Group 09**

*Siddharth Nagpal 2014B4A70821P*

*Amrit Kochar 2014B4A70821P*

**Problem Statement:** This research is aimed at the case of customers default payments in Taiwan and compare the predictive accuracy of default among did Data mining techniques.

We want to build a model which accurately predicts the probability of default payments by credit card customers.

**Dataset:** We are using an open dataset from UCI. [Link](https://archive.ics.uci.edu/ml/datasets/default+of+credit+card+clients)

**Attributes Information:** A binary variable, default payment (Yes = 1, No = 0) is the only response variable.

Here is the description of attributes in the selected dataset. ***(23 variables)***

X1: Amount of the given credit (NT dollar): it includes both the individual consumer credit and his/her family (supplementary) credit.   
X2: Gender (1 = male; 2 = female).   
X3: Education (1 = graduate school; 2 = university; 3 = high school; 4 = others).   
X4: Marital status (1 = married; 2 = single; 3 = others).   
X5: Age (year).   
X6 - X11: History of past payment. We tracked the past monthly payment records (from April to September, 2005) as follows:

*X6 = the repayment status in September, 2005;*

*X7 = the repayment status in August, 2005; . . .;*

*X11 = the repayment status in April, 2005. The measurement scale for the repayment status is: -1 = pay duly; 1 = payment delay for one month; 2 = payment delay for two months; . . .; 8 = payment delay for eight months; 9 = payment delay for nine months and above.*

X12-X17: Amount of bill statement (NT dollar).

*X12 = amount of bill statement in September, 2005;*

*X13 = amount of bill statement in August, 2005; . . .;*

*X17 = amount of bill statement in April, 2005.*

X18-X23: Amount of previous payment (NT dollar).

*(X18 = amount paid in September, 2005; X19 = amount paid in August, 2005; . . .*

*X23 = amount paid in April, 2005.)*

**Applications:**

* The credit to a customer by a business or a bank may be increased or the interest rate may be reduced in the case of improving creditworthiness of the individual or a business, while the opposite may occur if there is a decline in the customer’s credit profile. Therefore, this a major application for any customer-credit business.
* In a well developed and stable financial system, risk prediction plays an important role. The major purpose of risk prediction is to use financial information, such as business financial statement, customer transaction and repayment records, etc., to predict business performance or individual customer’s credit risk and to reduce the damage and uncertainty.

**Comment:** The estimated probability of default is a much more useful metric than classifying clients in credible or not credible.

**Methodology:**

First, we pre-process the data and remove the unwanted attributes. Since this is a binary class problem, we use SVM on the data and see the results. We will try out a few other classification methods as well if possible and chose the one which gives the best result.