Project proposal for Capstone 1

**Problem statement:**

Credit cards are the most common sources of finance which help people to realize their desires. At times, they also provide the much-needed support to tide over financial difficulties. It is important for every bank to know the credibility of its customer before issuing a credit card and assigning him/her the credit limit. To keep the integrity, a bank must avoid in investing the wrong customer who can default and cause loss to financial institutions. I thus propose to investigate customers are likely to default their next payment by applying Machine Learning models to the available data from Taiwanese Bank**.**

**Stakeholders:**

Banks will be interested in this solution as they may avoid losses by refusing to issue a credit card to the customer who is likely to default on their payments or lowering their credit limit.

**Dataset:**

The [dataset](https://archive.ics.uci.edu/ml/datasets/default+of+credit+card+clients) is available on UCI Machine learning repository. This dataset contains about 30,000 instances and 24 attributes. The datatype of all the attributes is int64.

|  |  |  |
| --- | --- | --- |
| **No** | **Attribute** | **Description** |
| 1 | LIMIT\_BAL | Amount of the given credit (NT dollar): it includes both the individual consumer credit and his/her family (supplementary) credit. |
| 2 | Gender | 1 = male; 2 = female |
| 3 | Education | 1 = graduate school; 2 = university; 3 = high school; 4 = others |
| 4 | Marital status | 1 = married; 2 = single; 3 = others |
| 5 | Age | Year |
| 6 | PAY\_1 – PAY\_6 | History of past payment from April to September, 2005  status is: -1/0 = pay duly; 1 = payment delay for one month; 2 = payment delay for two months; |
| 7 | BILL\_AMT1 - BILL\_AMT6 | Amount of bill statement (NT dollar) from April to September, 2005 |
| 8 | PAY\_AMT1 - PAY\_AMT6 | Amount of previous payment (NT dollar) ) from April to September, 2005 |
|  |  |  |

**Proposed Methodology:**

 Since we want to predict the likelihood of a defaulting the payment next month, a supervised classification algorithm is a perfect choice to build the predictive model

**Deliverable:**

The project would be delivered in the form of a video presentation with visualization and results. In addition, python jupyter notebook, report, publish an article in medium will give detail explanation on my approach.