Project 3 Proposal

Determining Credit Card Default with an Assortment of Supervised Classification Methods

Charlie Lew

Before any financial institution issues a credit card to a potential borrower it relies heavily on his or her credit history as a guide and to finally approve or disapprove the card. However, once the credit card is issued, it is difficult to determine ahead of time if the client will default. The current credit card default rate in the United States according to the Federal Reserve Bank of New York is at 5.4% for all borrowers but is at 8% for borrowers between ages of 18-29, which is rather high. Based on the purchase and payment history of the client, a bank could potentially predict if a borrower may default and thus take steps to shield itself from further risk in the future. The dataset is available from UCI Machine Learning Repository (https://archive.ics.uci.edu/ml/datasets/default+of+credit+card+clients)

The Minimum Viable Product would be to compared at least two supervised classification techniques to determine if a borrower will or will not default on a credit card payment. Metrics to look at will be turn-around-time and accuracy of the method.

The dataset contains recordings of 30,000 clients with 25 features. Table 1 below shows the features and value types for each

Feature	Type	Description
Customer ID	Integer	Anonymized customer ID
LIMIT_BAL	Integer	Amount of given credit
SEX	Integer	Gender, 1=male, 2=female
EDUCATION	Integer	Range from 1 to 6
MARRIAGE	Integer	Marital status, range from 1-3
AGE	Integer	Age in years
PAY_0	Integer	Repayment in September 2005
PAY_2	Integer	Repayment in August 2005
PAY_3	Integer	Repayment in July 2005
PAY_4	Integer	Repayment in June 2005
PAY_5	Integer	Repayment in May 2005
PAY_6	Integer	Repayment in April 2005
BILL_AMT1	Integer	Bill amount for September 2005
BILL_AMT2	Integer	Bill amount for August 2005

BILL_AMT3	Integer	Bill amount for July 2005
BILL_AMT4	Integer	Bill amount for June 2005
BILL_AMT5	Integer	Bill amount for May 2005
BILL_AMT6	Integer	Bill amount for April 2005
PAY_AMT1	Integer	Amount paid September 2005
PAY_AMT2	Integer	Amount paid August 2005
PAY_AMT3	Integer	Amount paid July 2005
PAY_AMT4	Integer	Amount paid June 2005
PAY_AMT5	Integer	Amount paid May 2005
PAY_AMT6	Integer	Amount paid April 2005
Default.payment.next.month	Integer	0=no, 1=yes, not used as a feature