**Domain**

Banking

**Dataset**

Loan Payment data

**Attributes**

* Loan\_id : A unique loan number assigned to each loan customers
* Loan\_status: Whether a loan is paid off, in collection, new customer yet to payoff, or paid off after the collection efforts
* Principal: Basic principal loan amount at the origination terms, could be weekly (7 days), biweekly, and monthly payoff schedule
* effective\_date: When the loan got originated and took effects
* due\_date: Since it’s one-time payoff schedule, each loan has one single due date
* paidoff\_time: The actual time a customer pays off the loan
* pastdue\_days: How many days a loan has been past due
* age, education, Gender: A customer’s basic demographic information

**Problems**

The Loan Payment dataset contains above mentioned attributes. We would like to perform classification on Loan\_status variable that contains three categories namely PAIDOFF,COLLETION AND COLLECTION\_PAIDOFF. The PAIDOFF category is all about loan paid off by the customers. The COLLECTION is all about the loan is yet to be paid off so bank is collecting money from those customers who seem to be unwilling to repay the loan amount; these customers are apprehended to be defaulters.

The next and last category is COLLECTION\_PAIDOFF that states that the loan amount was repaid by the customers after conducting Collection drive.

The objective is to **Classify customers into any of the given categories.**

This analysis will generate valuable insights for banks to consider which customers they will provide loan to.