

# Liability account monitoring for Money Mules

IIT Bombay case study

## Introduction

Bank A allows customers to open savings accounts through online channels. Customers availing this journey can execute all the steps (including providing details and documentation, KYC and account funding) online without the need to visit a physical branch or meeting a Bank employee face to face. Introduction of this journey has led to significant improvement in customer ease and experience.

However, some unscrupulous elements have been found to be taking advantage of this facility. Such individuals (henceforth referred to as mules) open savings accounts with Bank A with the sole purpose of defrauding others. Mules use various social engineering techniques to fraudulently convince unsuspecting individuals to transfer money from their Bank accounts to the accounts that the mules have opened with Bank A.

## Problem statement

To respond to illegal usage of its facilities, Bank A has decided to proactively monitor transactions done by customers through their Savings Accounts. To achieve this, Bank A needs to develop an automated transaction monitoring framework using Machine Learning algorithms.

Your objective is to develop this framework.

## Datasets

You have been provided with a random sample of 1,00,000 savings account details in “Dev\_data\_to\_be\_shared.zip”, along with a flag (target) – henceforth known as “development data”. Accounts that have been identified as mules have target = 1. You have also been provided with several independent variables. These include:

- Account level attributes like account opening date
- Demographic attributes of the customer (e.g. income, occupation, city tier etc.)
- Transaction history of customer (includes credits / debits of specific types over a period of time)
- Other attributes like product holding with the Bank, app / web logins etc.

Some variable names have been provided. For other variables, demographic attributes start with “demog\_”, transaction attributes start with “txn\_” and the rest start with “others\_”,

You have also been provided with another random sample of 50,000 savings account details in “validation\_data\_to\_be\_shared.zip” with the same set of input variables, but without “target”. This will be referred to going forward as “validation data”.

## Requirements

Using the data provided, you will have to come up with a way to predict the probability that a given account is a mule. You can use the development data for this purpose.

You are then required to use the same logic to predict the probability of all the accounts which are a part of the validation data. Your submission should contain two columns – the Primary key from the validation data, and the predicted probability against that account.

You are also required to submit a detailed documentation of this exercise. A good document should contain details about your approach. In this section, you should include a write up on any algorithms that you use. You should then cover each of the steps that you have followed in as much detail as you can. You should then move on to any key insights or observations that you have come across in the data provided to you. Finally, you should write about what metrics you have used to measure the effectiveness of the approach that you have followed.

## Evaluation

As detailed in the previous section, you are required to submit the Primary key and predicted probabilities of all the accounts provided to you in the validation data, as well as a documentation.

We will only evaluate submissions that are complete and pass sanity checks (probability values should be between 0 and 1 for example).

Submissions will be evaluated basis how close the predicted probabilities are to the actual outcome. We will also evaluate the documentation basis it's completeness and accuracy. Extra points will be granted to submissions that include interesting insights / observations on the data provided.