

# 第一次编程作业：银行客户流失预测

## 任务背景：

每家银行都希望能留住客户以维持其业务。某银行整理了他们的客户数据，希望你能帮助预测客户是否流失（Churn）。

## 数据集：

数据为附件中提供的 `supply_chain_train.csv` 文件(需要在此文件的数据集内，自行划分训练集和测试集，而非使用原网站的测试集)。其中最后一列（`Attrition_Flag`）列为标签。

数据集具体说明可见以下网站：

<https://www.kaggle.com/datasets/celocruz/bank-database>

## 任务说明：

请实现以下任务并在报告中详细记录所采用的解决方案

1. 实现逻辑回归算法，学习一个二分类器，以预测用户是否会流失。
2. 需要数据预处理，如数据清洗、归一化、非数值数据处理等
3. 需要自行划分训练集和验证集。分别使用 `precision`, `recall`, `accuracy` 作为评价指标，并判断是否欠拟合、过拟合。将训练过程的损失函数以及指标的变化用曲线可视化。
4. 尝试将该问题建模成一个回归模型，使用 `MSE`（均方误差）作为损失函数来进行训练，使用 `precision`, `recall`, `accuracy` 作为评价指标。将训练过程的损失函数及指标的变化用曲线可视化(推荐使用 `TensorBoard` 或 `Wandb` 等可视化工具)。对比分类模型和回归模型的异同，并分析背后的原因。

## 要求：

1. 使用 `python` 编程语言
2. 自行实现算法。请勿直接调用高度封装的工具包(如 `sklearn` 中的 `model.fit()` 函数) 直接训练模型；但可在每一步迭代中，调用线程的损失函数及梯度下降方法（如 `pytorch` 中的 `loss.backward()` 函数）更新模型
3. 提交实验报告和代码文件。需要包含任务说明中各项的完成情况、损失函数公示推导、预处理情况、算法具体实现方式、对结果的分析。代码需要有关键注释。