

## 实验 6：银行潜在客户挖掘实验指南

### 一、实验目的

本实验旨在对金融用户数据进行数据挖掘与分析，从中找出目标客户，并对结果进行评估分析。具体实验内容为根据葡萄牙某银行机构获得的用户数据，结合相应的标签（标签内容  $y$  为向该用户进行的推销是否成功）进行数据挖掘，判断用户是否为潜在客户，并评估预测结果。

### 二、实验步骤

#### 1. 数据获取

数据说明与下载地址（需注册 Kaggle 账号后可以下载数据）：

<https://www.kaggle.com/henriqueyamahata/bank-marketing/downloads/bank-marketing.zip/1>

#### 2. 数据清洗

实现代码中的 `data_preprocess()` 函数：

输入：数据清洗之前的训练数据  $X$

输出：数据清洗之后的训练数据  $X$

#### 3. 训练/测试集划分

（已实现）`split_data()` 函数：

输出：

$x_{train}$ : 训练集中的输入特征

$y_{train}$ : 训练集中的输出特征

$x_{test}$ : 测试集中的输入特征

$y_{test}$ : 测试集中的输出特征

#### 4. 模型训练

实现 `predict()` 函数：

输入：

$x_{train}$ : 训练集中的输入特征

$y_{train}$ : 训练集中的输出特征

$x_{test}$ : 测试集中的输入特征

输出：

$y_{pred}$ : 标签预测列表（1 表示预测是潜在客户，0 表示预测不是潜在客户）

注意：

1. 可以利用实验 3 中实现的实现 LR、SVM 和感知机 (Perceptron) 三种分类算法的其中一种。
2. 可以利用 sklearn 中自带的分类器进行分类, 详细的算法参考 FinMKT.py 中的#some usable model
3. 鼓励对比不同的分类算分对结果的影响。

## 5. 数据分析 (附加)

附加题目, 不强制要求做。鼓励对用户数据进行可视化分析, 具体可以参考:

[https://blog.csdn.net/qq\\_35654046/article/details/85234958](https://blog.csdn.net/qq_35654046/article/details/85234958)

## 三、提交内容

### 1. 实验报告:

1. 通过实验了解金融营销的应用场景;
2. 对比金融营销场景与其他金融科技场景算法的不同。
3. 通过对银行潜客挖掘数据的分析, 多维地分析潜在客户的特征; (附加)

### 2. 实验代码:

补充完整的 FinMKT.py 文件