# 基于\*\*的客户流失预测模型研究

网络金融部/创新应用科 王许一 李飏

**内容摘要：**

本文针对竞赛数据给出的电信客户相关信息，使用\*\*方法，进行客户流失预测建模研究，通过\*\*\*过程，取得\*\*成果，同时结合数据分析下探，根据客户在公司停留的生命周期阶段，按事前干预、事中运营和事后挽回，给出三种层次的分类营销建议。

## 题目解析与研究目标

### 题目解析

本竞赛课题为客户流失预测，数据集为电信客户的简单身份信息（性别、是否老年人、是否有家属、停留月数等）、签约过的服务信息（签约的具体服务、合同期限、付费方式等）、消费过的金额信息（每月消费金额、累计消费金额等），根据数据集采用相关模型方法、数据分析等手段，预测客户的流失，同时提出相应的挽回方案。

目前各行各业对客户竞争越来越激烈，获取新用户的成本越来越高，如何在维护现有用户情况下进行拓新，成为通信公司维持用户数、保证收入的一个重要因素。

本竞赛课题为电信客户流失预测，数据集包含6043条客户记录，21维特征，主要分为如下方面：

基本属性：年龄、性别、是否有亲属、是否有伴侣；

签约信息：电话服务、多线服务、因特网服务、在线安全、在线备份、 设备保护、技术支持、流媒体电视、流媒体电影；

账户信息：客户ID、在网时长、合同方式、无纸化账单、付款方式、月消费、总消费；

流失特征：是否流失；

本文将通过上述已有特征，通过数据分析探索衍生相关特征，采用机器学习方法，并结合商业目标调整预测结果，实现收益最大化。

### 研究目标

本文根据竞赛课题要求进行解析，旨在分析出一套最大化覆盖流失客户的模型方案，再结合相关数据分析，按客户在公司的生命周期时间轴，以事前干预、事中运营和事后挽回三个层次，提出一定的营销建议。

## 研究思路和方法

### 1、建模目标

本文的建模分析思路主要是使用大数据模型方法，从现象发现结论，实现流失客户名单，并根据数据分析结果给出挽留及促活方案。

### 2、建模流程

本文的数据分析思路是从已有标签的数据结论出发，总结分析相关的特征，辅助相关的营销分析。

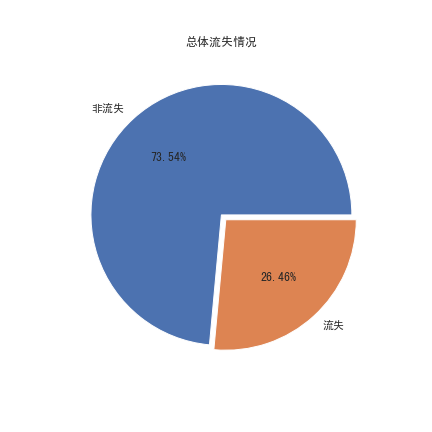
2.1 数据清洗

通过数据分析探索发现，总付费金额类型为object，应转为float，继而进行缺失值统计，发现有总消费有9个缺失，同时均为在网时长为0。

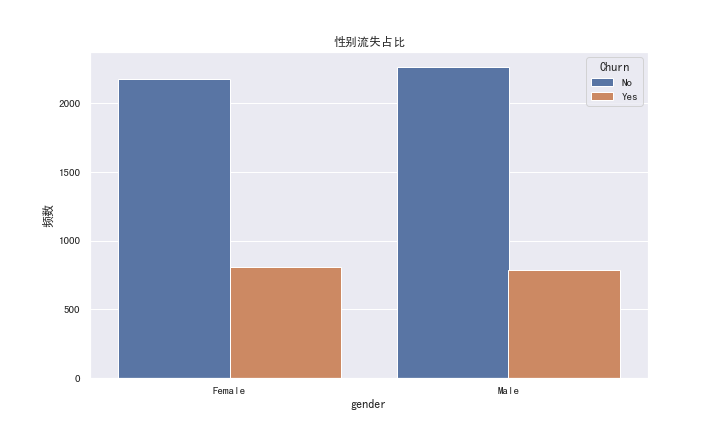
2.2 数据探索

将从客户属性、订阅服务、账户信息三个方面进行数据探索。

总体流失情况



2.2.1 客户属性

a.性别流失

b.

2.3 特征工程

2.4 模型选择

经过前期探索，最终选用传统逻辑回归Logisticregression、集成模型XGBoost、支持向量机SVM作为主要预测模型。

数据探索中看出数据存在不均衡现象，结合商业目标，以减少客户流失为目标，最大限度识别流失客户，在进行建模过程中将对目标变量权重进行加权，提高流失客户权重，降低存留客户权重

选择常用机器学习

2.5 模型评估

通过验证集对上述模型进行最终评估，以F1-Score、AUC-Scrore、Recall-Score三种评价指标为依据，结果如下：

三、结论