2022 年度“杉数杯”数学建模精英联赛

C 题：云服务用户的流失分析

公有云服务指为用户提供可通过互联网访问的虚拟服务器空间及其配套资源。在公有云中，所有硬件、软件和其他支持性基础结构均为云服务商拥有和管理，用户可以按需购买云服务器、数据存储和其他云相关服务并通过互联网访问这些服务器。相较于构建一个传统数据中心，公有云不仅可快速提供服务节省时间成本，并且可为用户节省购买、管理和维护本地硬件及应用程序基础结构的昂贵成本。此外，公有云也具有较强的弹性服务的能力。以微博为例，尽管其功能众多，但冗余较少，当出现突发热点事件时，会面临核心服务的流量数倍增长，这就导致弹性扩容需求的产生。由于公有云服务具有较好的弹性伸缩能力，因此微博的可伸缩业务如果利用公有云进行弹性部署，将完美解决突发热点事件导致流量激增的需求。目前，公有云通常用于提供网上办公应用，机器学习的训练、下载及存储，游戏开发和环境测试等。

近年来，随着公有云业务的广泛普及与互联网市场占比的日趋饱和，云服务厂商的业务发展方向也逐渐从大力扩张转为优质服务。某云服务厂商通过对其用户（含个人用户及企业用户）使用云资源的各项指标进行观测，发现有部分用户在使用了一段时间云服务后不再继续使用，它将这类客户定义为“流失用户”（附件 1 中提供了 250 名流失用户的样例）。由于用户的实际使用和自身业务的特点，监控指标发生突变较为正常，因此当监控指标发生长期的趋势性变化时，才能判定为“流失用户”。该云服务厂商对流失用户进行问询或挽留将有助于提高云服务质量，规避用户流失风险和提高云服务商发展潜力。

因此，该云服务商聘请你们的队伍担任企业顾问，根据用户使用云资源的监控数据，分析使用规律，构建用户流失预警模型，希望在某用户彻底流失之前，根据它的自身属性及行为等特征识别出该用户的流失风险。以此来帮助云服务厂商有效识别有离网倾向的用户，为该云服务厂商带来可观效益。

请你们团队结合实际情况，对相关数据进行深入分析，研究下列问题:

(1) 根据附件 1 中的流失用户监控指标的监控值，建立筛选指标模型，选出你们队伍认为与用户流失相关的重要指标，请说明选取的指标数量以及原因。

(2) 根据附件 1 和附件 3 中的用户资源利用情况，建模刻画用户画像，对用户的流失风险进行分级，给出每一流失风险等级用户特征的数学描述。为判断用户流失风险提供定量参考。

(3) 基于问题（1）筛选出的重要监控指标，根据附件 1 与附件 3 中的用户监控指标的监控值，构建用户流失预测模型，说明流失用户的具体判别标准，特别是流失用户的监控指标的长期变化趋势特征。利用附件 1 中用户监控指标的监控值，计算该模型的精确率、召回率及 F1-score 来评价该模型准确性，并分析模型对相关因素的依赖性。根据附件 4 中用户监控指标的监控值，利用该用户流失预测模型，预测附件中的流失用户，并将流失用户 ID 填入附件 5 中。其中：

TP：预测为 1，实际为 1，预测正确。
FP：预测为 1，实际为 0，预测错误。
FN：预测为 0，实际为 1，预测错误。
TN：预测为 0，实际为 0，预测正确。

- 精确率(Precision)： 是针对预测结果而言的，是在被所有预测为正的样本中实际为正样本的概率，公式为$\text{precision} = \frac{TP}{TP+FP}$
- 召回率(Recall)： 是针对原样本而言的，是在实际为正的样本中被预测为正样本的概率，公式为$\text{recall} = \frac{TP}{TP+FN}$
- F1-score 表达式为：$F_1 = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recal}}$

(4) 根据附件 4 中用户监控指标的监控值，结合问题（2）构建的模型，预测用户的最终流失时间点所在范围，并把对应选项字母 填入附件 5 中。

时间点范围选项：

A：1 个月以内 B:1 到 3 个月 C：3 到 6 个月 D： 6 个月以上

注 1: 请将问题 3 中的流失用户 ID、最终流失时间点和问题 4 中的流失预警时间点填入附件 5 中，并作为支撑材料（勿改变文件名）跟论文一起提交。

注 2: 由于云服务商无法知道用户使用云的具体业务，因此只能监控主机的物理性能或者网络性能，即监控指标为该云服务厂商监控到的用户对云服务产品（计算、存储、网络）的使用情况，共有 156 个监控指标。

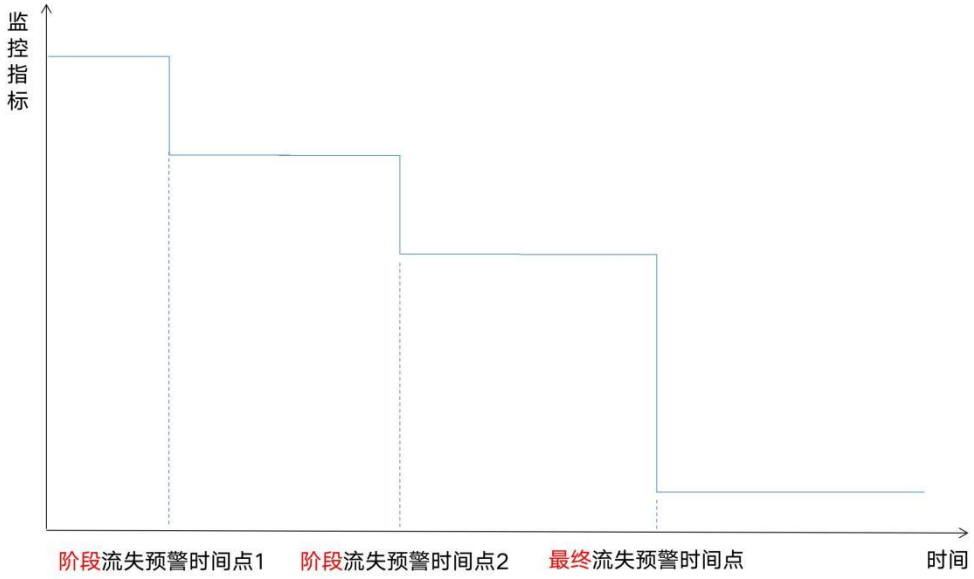
注 3：最终流失时间点为在该点之后，用户监控指标维持低值或趋近于 0 且不再回升，用户已彻底流失。

注 4：在本题中不进行对阶段流失预警时间点的考量，因此参赛选手并不需要使用附件 2 中的数值，但为了更完整地介绍本题场景，因此对附件 2 中的流失预警时间进行以下的补充说明：

1.已知根据监控指标预测出的某些用户的流失预警时间会早于附件 2 中给出的对应流失预警时间点，并且每个流失预警时间点需要依据该点之前的数据进行预测。

2.阶段流失预警时间点为在用户彻底流失之前，监控指标大幅下降的时间点。部分用户可能没有阶段流失时间点（见附件 2）。

阶段流失预警时间点及最终流失预警时间点如下图所示：



附件列表：

附件 1：一年内 250 名流失用户监控值

附件 2：流失预警时间表

附件 3：一年内 182 名正常用户监控值

附件 4：一年内 200 名未检测混合用户监控值

附件 5：预测结果表

参赛选手必须保证以上附件中所有数据的安全，不能进一步进行扩散，并且在该比赛结束之后，参赛选手需要确保将数据及时删除。