# 问题分析

## 问题描述

Click-Through Rate Prediction

Predict whether a mobile ad will be clicked

In online advertising, click-through rate (CTR) is a very important metric for evaluating ad performance. As a result, click prediction systems are essential and widely used for sponsored search and real-time bidding.

For this competition, we have provided 11 days worth of Avazu data to build and test prediction models. Can you find a strategy that beats standard classification algorithms? The winning models from this competition will be released under an open-source license.

## 评价指标

Submissions are evaluated using the Logarithmic Loss (smaller is better).

-log P(yt|yp) = -(yt log(yp) + (1 - yt) log(1 - yp)), yt为真实值, yp为预测概率

## 提交文件格式

最终输出的'click'为点击的 概率

Submission Format The submissions should contain the predicted probability of click for each ad impression in the test set using the following format:

id,click

60000000,0.384

63895816,0.5919

759281658,0.1934

895936184,0.9572

......

## 问题分析

目标为'click', 取值为0,1, 是典型的二分类问题

常见的分类及CTR模型:

* 朴素贝叶斯(Naive Bayes, NB)
* Logistic回归(Logistic Regression, LR)
* 决策树(Decision Tree, DT)
* 梯度提升决策树(Gradient Boosting Decision Tree, GBDT)
* 随机森林(Random Forest, RF)
* 支持向量机(Support Vector Machine, SVM)
* 优化后的LR+L1(Follow-the-regularized-Leader, FTRL)
* 因子分解机FM(Factorization Machine, FM)
* 场感知因子分解机(Field-aware Factorization Machine, FFM)