**音乐网站用户流失预测项目：特征工程总结及解析**

1. **模型采用协同过滤算法：**协同过滤算法由用户、物品、打分三个特征来进行计算，在特征探索时发现train数据集没有指定打分的特征，所以需要自己构建一个打分标准，在进行特征工程时，我们可以将target续订作为打分计算的基础，具体做法为：1、计算用户对电影的累计续订值 2、根据每个电影的target值除以用户的累计续订值得到一个概率性分数，可以以此作为R矩阵中的打分值。同时，我们也要对数据中的异常值和缺失值进行处理，异常可以直接剔除，缺失值可以采用mean填充、如果同行特征都缺失较多可以将该样本剔除。

2、**采用ctr（分类或回归）算法（比如lightgbm）：**在进行ctr算法特征工程时，考虑train数据集基本为类别型特征（object），我们可以用MeanEncoder进行类别型特征处理，然后将所需其他特征进行融合（songs.csv、members.csv、song\_extra\_infos.csv等其他数据集特征进行融合），，融合时考虑具体模型需要进行具体处理，然后一起送入模型进行训练。