# Week3 Thinking

### Thinking1 如何使用用户标签来指导业务（如何提升业务）

根据用户标签可以获取用户的一些特征，用户画像，精准营销，获取客户；根据我们的item特征，与用户特征做一些关联，推荐用户感兴趣的item,粘住客户；最后，也要做一些预测，比如说有一些客户可能会流失，及时采取一些措施留住客户。

### Thinking2 如果给你一堆用户数据，没有打标签。你该如何处理（如何打标签）

1. 可以让机器去打标签，先跑出几类，再去看下这每个类别到底代表什么含义，通过后面再做具体的分析，给类别一个有含义的标识。  
2. 看当下热门的标签是否符合

### Thinking3 准确率和精确率有何不同（评估指标）

精确率考虑的是预测结果为正的情况下的准确率  
准确率考虑的维度更多，不光考虑预测结果为正的，还有预测结果为负的时候，只有预测结果是对的都会考虑。

### Thinking4 如果你使用大众点评，想要给某个餐厅打标签。这时系统可以自动提示一些标签，你会如何设计（标签推荐）

1. 直接使用系统最热门的标签  
2. 使用这家餐厅里相应物品上最热门的标签  
3. 使用用户经常使用的标签  
将2和3进行加权融合，生成最终的标签推荐结果

### Thinking5 使用TPOT等AutoML工具，有怎样的好处和不足？

好处：

1. 可以解决

1）特征选择（基于树模型、基于方差、基于F-值的百分比）

2）模型选择  
 3）数据预处理（二值化、聚类、降维、标准化、正则化等）  
2. 处理小规模数据非常快

3.支持的分类器和回归器很多，有：

支持的分类器主要有贝叶斯、决策树、集成树、SVM、KNN、线性模型、xgboost

支持的回归器主要有决策树、集成树、线性模型、xgboost

4.可以实现通过代码来写代码，可以说是机器学习的机器学习，可以通过export()方法把训练过程导出为形式为sklearn pipeline的.py文件

不足：  
1.目前只能做有监督学习

2. 可以做特征模型选择，但不包括数据清洗  
3.处理小规模数据非常快，但大规模数据非常慢。可以先抽样小部分，使用TPOT