构建多分类模型，预测出用户所属的标签分类。  
32万左右用户标注信息作为训练集，8万左右数据作为预测集。  
结果：预测集数据的预测结果的F1平方值。  
模型2级分类器，主要用Lgb  
我已经做了一个版本f1是90左右，但是数据不平衡，正样本多，负样本少，需要更精确挖掘小的负样本分类  
数据、特征、模型都可以改，由于训练集测试集分布不一致，导致需要改进。

数据增强

特征选择或者正则化

模型loss：cross entropy？

优化器选择：adagrad适用于稀疏样本学习？

Adam？？