1. 赛题为 新闻文本分类
2. 赛题数据

样本的标签有14个类别，分别为：财经、彩票、房产、股票、家居、教育、科技、社会、时尚、时政、体育、星座、游戏、娱乐。

样本按照字符级别进行了匿名处理，训练集20w条样本，测试集A包括5w条样本，测试集B包括5w条样本。

1. 评价标准

评价标准为类别f1\_score的均值

1. 思路
2. **one-hot/tf-idf** 构建词特征向量，用**机器学习**方法（LR，SVM等）做出分类
3. **word2vec**构建词特征向量，结合**深度学习**方法（textcnn，LSTM，textR-cnn,attention等）做出分类
4. 结合**bert**做出分类