

NLP with Disaster Tweets

模型部分

分别去调研 EDA、Xgboost、Ensemble

所有人都要看BERT，起码知道怎么搭模型

要求对调研的模型，阅读相关 paper 或者 blog，可以去详细解释模型的基本原理。要整理成文档。

代码部分

从[别人的 submission](#)中寻找代码

1. 这个代码是如何做特征工程的，如何分析数据的统计学信息、如何构建新特征、如何处理缺失值、如何可视化等
2. 了解这个代码是怎么调参的、模型是怎么搭起来的，最次要能做一个掉包调参侠，最好则是了解模型原理可以手写模型

其他参考博客和代码

[BERT](#)

[Sklearn](#)

[BERT](#)