特征工程的方法整理：

<https://www.cnblogs.com/jasonfreak/p/5448385.html>

归一化、三个特征选择法（简单）、降维调包

**生成尽量多的 Feature，相信 Model 能够挑出最有用的 Feature、Kaggle 上最常用的模型基本都是基于树的模型：**

**Gradient Boosting、Random Forest、Extra Randomized Trees**

将训练数据的一部分划出来作为验证集

给新手的一些建议：

选择一个感兴趣的比赛。如果你对相关领域原本就有一些洞见那就更理想了。

根据我描述的方法开始探索、理解数据并进行建模。

通过 Forum 和 Scripts 学习其他人对数据的理解和构建 Feature 的方式。

如果之前有过类似的比赛，可以去找当时获奖者的 Interview 和 Blog Post 作为参考，往往很有用。

在得到一个比较不错的 LB 分数（比如已经接近前 10%）以后可以开始尝试做 Ensemble。

如果觉得自己有希望拿到奖金，开始找人组队吧！

到比赛结束为止要绷紧一口气不能断，尽量每天做一些新尝试。

比赛结束后学习排名靠前的队伍的方法，思考自己这次比赛中的不足和发现的问题，可能的话再花点时间将学到的新东西用实验进行确认，为下一次比赛做准备。

<https://www.kaggle.com/benhamner/python-data-visualizations>

特征画图

apachen强调协方差比较特征