# 执行力

# 如何与客户沟通

# 书到用时方恨少

学的时候总抱怨方法、原理太多，真正用的时候发现一个也用不上。学的时候总感觉能做好大的项目，用的时候一个特征汇总就搞掉你。

介绍一些架构。

# 不平衡分类

# 有限总体稀有事件抽样

# 做事要做漂亮

在规定时间内做事，宁愿做的不要那么多也要那么精。

# 数据归一化对算法的影响

1. 一些算法天然的要求数据量纲一致：kmeans、knn、pca等
2. 除此之外，数据归一化还能够影响梯度下降等最优化方法的收敛。如神经网络在进行训练之前经常需要对量纲差异较大的数据进行归一化。

# 做一些项目的基本点

1）产品设计

对分布式来说，产品设计往往是其短板，由于其数据处理过程都是基于rdd的因此，在中间交互、结果展示等方面都受限

2）数据处理

3）展示

# 对于不定参数的对象

在应用中有些参数需要不定长的参数,比如有的需要列名/时间/分箱,有的需要列名/时间就可以.

此时两种可以解决上述需求:

1)将类写为class A(private var 参数1: T){以set形式传递参数}的形式

2)类正常写为class A(val 参数1: Option[T]）,以Option包括参数，直接apply

# 常用的工程项目目录结构和理念

## 理念

低内聚高耦合

## 目录方式1

Utils

Test

Models

Objects