# **【总】特征输入&特征工程**

@王中元

# Summary

基于depth、trade、order等数据生成可能对日频级别预测有效的输入特征，并进行相应的特征工程，提升模型表现上限

# 任务拆分

## **特征生成**

* 通过逻辑生成特征并聚合到分钟、10分钟、30分钟等频率，需考虑到其可能的（相比中高频）更长期预测能力
* 通过数据驱动生成特征，例如特征交叉、遗传算法等，其中需考虑到评估目标与复杂度等

## **特征工程**

* 根据特征数据分布与启发性先验，进行特征以预处理（如各种类型的归一化，填充，异常值处理）
* 如果生成特征过多，需要进行特征筛选，除基于逻辑的筛选外具体可参考

[请至钉钉文档查看附件《因子筛选与特征工程调研.pdf》](https://alidocs.dingtalk.com/document/edit?dentryKey=anZvVrXlH91Dgn4G&iframeQuery=anchorId%253DX02l9nxv3vkq187yyk83ng&type=d&utm_medium=drive_spacefile&utm_source=drive)