



赛题：基金产品的长周期申购和赎回预测

队名：CEDC冲冲冲

# Contents 目录 ▶

01

团队介绍

02

方案综述

03

解决方案

04

方案成效

05

技术创新



# 01 团队介绍



**队长:毛锋**

中国电子信息产业集团有限公司

博士, 清华大学控制科学与工  
程专业

中国电子专员



**队员1:林丰玉**

中电数据产业集团

硕士, 西安交通大学数学  
与统计专业

数据产业集团技术研究员



**队员2:黄颖坤**

中电数据产业集团

博士, 西南交通大学电气  
专业

数据产业集团技术研究员



# 02 方案综述

## 任务目标

根据20只基金的历史申购和赎回时间序列以及有限个平台侧关键特征，借助大模型自行获取和构造其他有效特征，训练时序模型预测每只基金在（2025/7/25-7/31）7天内每天的申购量和赎回量。

## 赛题挑战

**难点1** 基金极端申购、赎回数据的处理和预测。

**难点2** 如何使用大模型获取和构造有效特征，并保证稳定性。

## 赛题数据

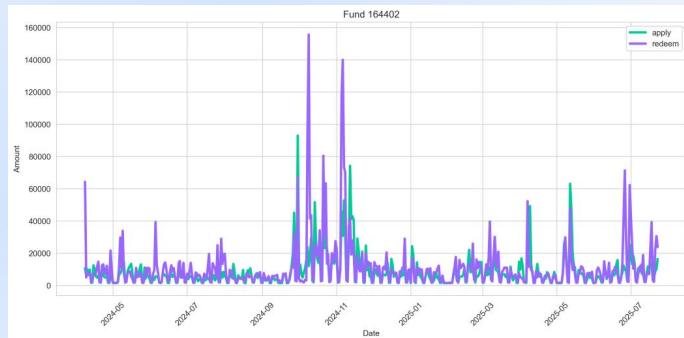
赛题数据

基金code

交易日期

申购金额、赎回金额

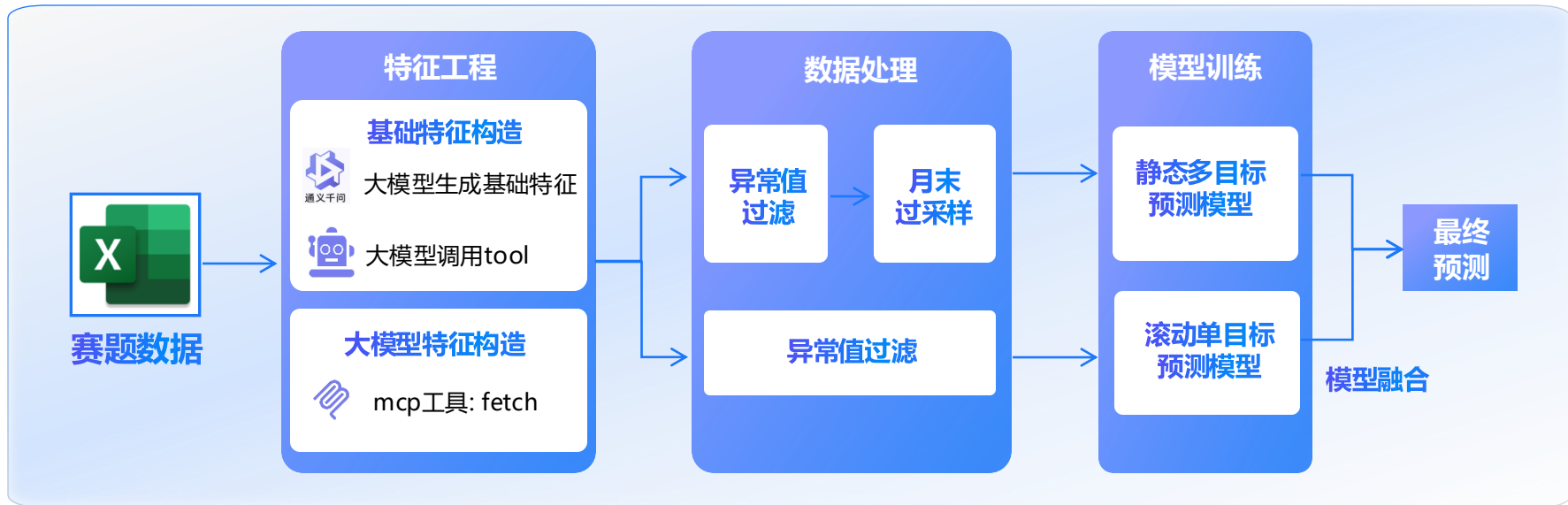
申赎核心页面1/2/3曝光uv



## 技术方案

**创新点1** 大模型驱动的构建基金预测模型框架。

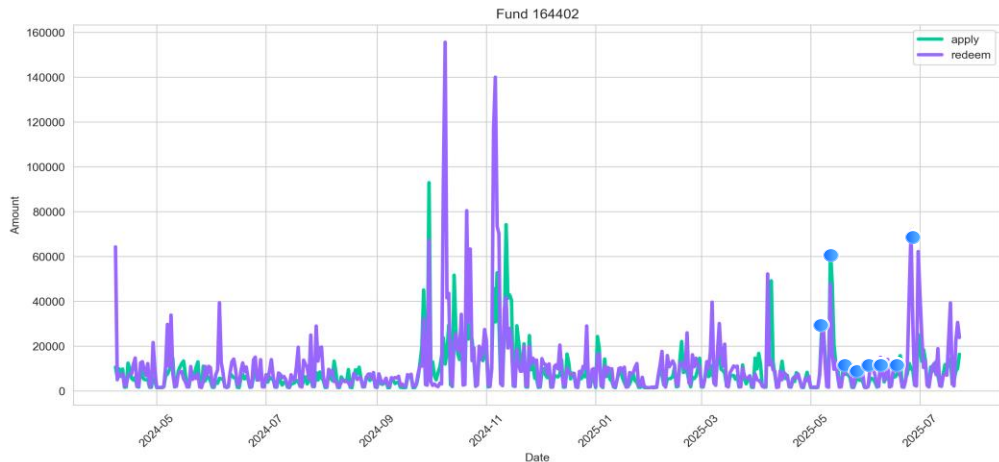
**创新点2** 大模型特征构造模块可拓展性强。





# 03 解决方案

## 申购、赎回量分析



**现象1** 大部分情况下，各支基金的申购、赎回量具有1周、1月的周期性。

**现象2** 各支基金的申购、赎回量存在：工作日>周六>周日。节假日明显减小。

```
prompt_round1 = f"""
```

根据上述文件，目标是预测各支基金未来一周的申购和赎回量。基金的购买和赎回具有1周，1月的周期性。在特征设计中，使得预测结果，周末比工作日小，周日比周六小，且节假日申购/赎回会减小。如何设计特征？

```
"""
```

```
prompt_round2 = f"""
```

根据上述特征工程设计方案，生成python代码，使用Lightgbm模型预测未来一周各支基金每天的申购和赎回量。

```
"""
```

## 基础特征构造

### 大模型计算工具使用



计算基础特征

### 基础特征构造

#### 时间特征

基础时间特征: dayofweek、dayofmonth、dayofyear、weekofyear

时间周期编码特征: dayofweek\_sin/cos、dayofmonth\_sin/cos、dayofyear\_sin/cos、weekofyear\_sin/cos

时间标记特征(工作日、周末、节假日、月初/末、季初/末、月初第1个工作日、月末最后1个工作日)

#### 申购/赎回特征

滚动特征: 不同窗口(3/7/14/30天)申购和赎回量的统计量(均值、标准差及波动率)

滞后特征: 过去特定天数(3/7/14/30天)的申购和赎回量

#### 页面曝光uv特征

滚动特征: 不同窗口(3/7/14/30天)页面曝光uv的统计量(均值)

#### 申购/赎回统计特征

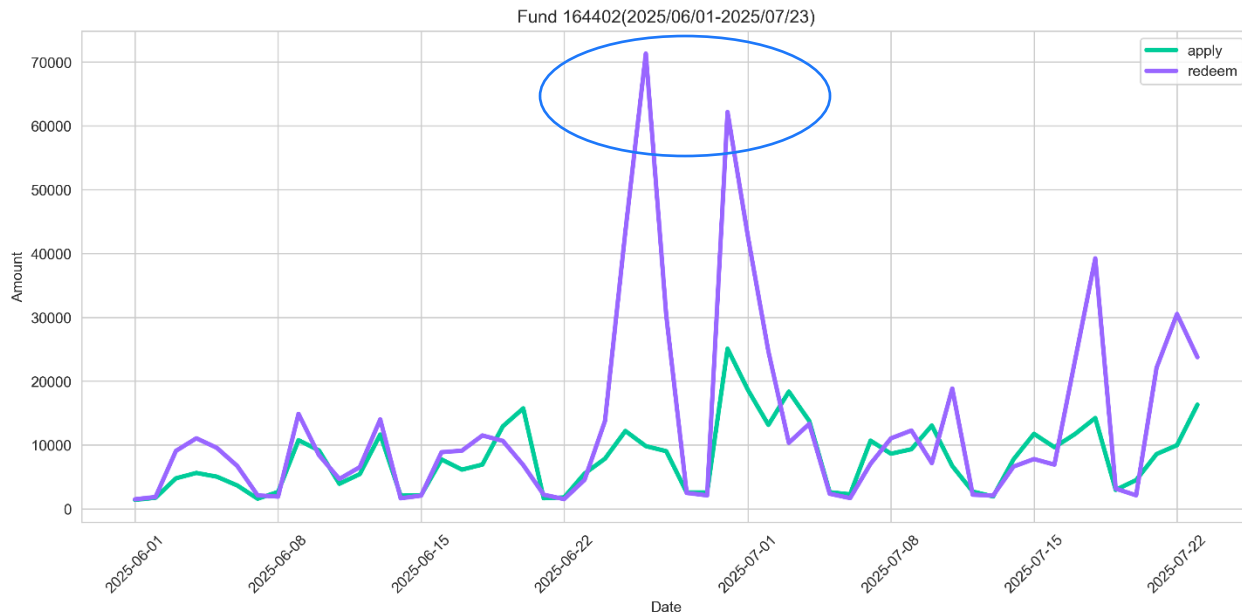
工作日和周末申购和赎回量的统计量(均值、标准差), 周末调整系数

当月申购和赎回量的统计量(均值、标准差)

月初(前3天)和月末(28号后)的申购和赎回的统计量(均值、标准差)

## 挑战1：极端数据处理与预测

现象3 部分基金月底最后1个工作日会大量赎回

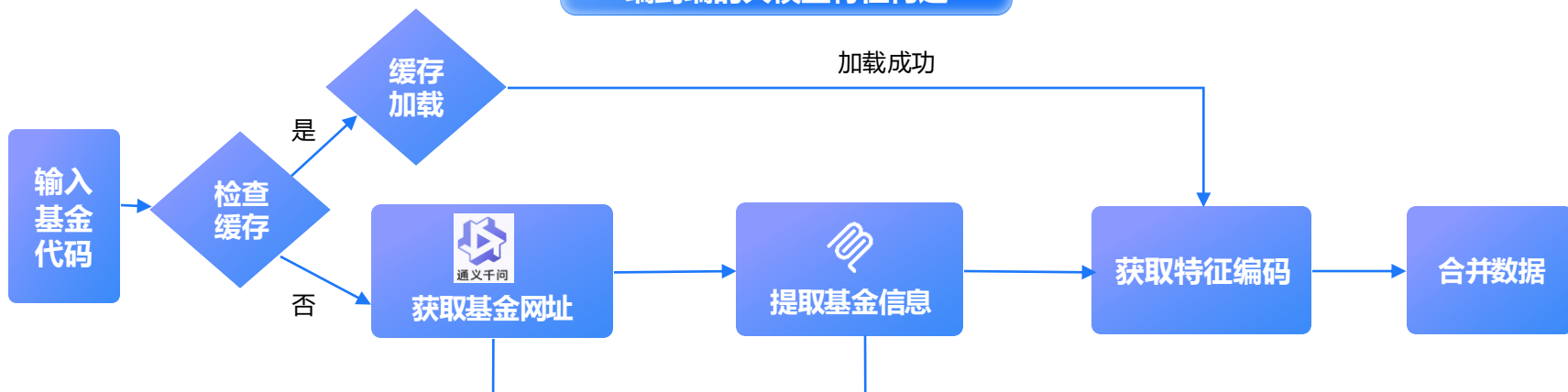


## 挑战2：大模型特征构造

基金代码	基金名称	基金类型	基金代码	基金名称	基金类型
000086	南方稳利1年持有期债券型证券投资基金	债券型	050027	博时信用债纯债债券型证券投资基金	债券型
000192	富国信用债债券型证券投资基金	债券型	100050	富国全球债券证券投资基金(QDII)	债券型
000385	景顺长城景颐双利债券型证券投资基金	债券型	160717	嘉实恒生中国企业指数证券投资基金(QDII-LOF)	QDII
000402	工银瑞信纯债债券型证券投资基金	债券型	161024	富国中证军工指数型证券投资基金(LOF)	股票型
000965	汇丰晋信新动力混合型证券投资基金	混合型	161028	富国中证新能源汽车指数型证券投资基金(LOF)	股票型
001016	华夏沪深300指数增强型证券投资基金	股票型	164402	前海开源中航军工指数型证券投资基金	股票型
001316	安信稳健增值灵活配置混合型证券投资基金	混合型	164906	交银施罗德中证海外中国互联网指数型证券投资基金(LOF)	QDII
001338	安信稳健增值灵活配置混合型证券投资基金	混合型	166016	中欧纯债债券型证券投资基金(LOF)	债券型
001407	景顺长城稳健回报灵活配置混合型证券投资基金	混合型	168204	国联中证煤炭指数型证券投资基金	股票型
001480	财通成长优选混合型证券投资基金	混合型	530028	建信短债债券型证券投资基金	债券型

## 挑战2：大模型特征构造

### 端到端的大模型特征构建



`prompt = f"""你是一名专业的金融分析师，请提供基金集合{fund_codes}的天天基金基金概况网址。  
输出格式-JSON对象,示例:  
{{"fund_code": "000086", "fund_url": "https://fundf10.eastmoney.com/jbgk_000086.html"}, ...}"""`

`mcp_fetch_prompt = f"""请从下面网址提取基金的结构化描述. 包括以下字段:`

- 基金类型 ('债券型': 1, '混合型': 2, '股票型': 3, 'QDII': 4)
  - 基金规模 ('小型基金(资产规模<10亿元)': 1, '中型基金(资产规模10-50亿元)': 2, '大型基金(资产规模50-100亿元)': 3, '超大型基金(资产规模100亿元以上)': 4)
- 输出格式-JSON对象, 示例: `{{ "fund_code": "{fund_code}", "fund_type": 1, "fund_size": 2 }}`, 网址: `{fund_url}` """

## 挑战1：极端数据处理与预测

**样本过滤** 去除z-score > 3的样本数据

$$z = \frac{X - \mu}{\sigma}$$

其中:

z —— z-score值

X —— 原始样本申购/赎回量值

$\mu$  —— 基金申购/赎回量均值

$\sigma$  —— 基金申购/赎回量标准差

**过采样** 对月末3天数据进行3倍过采样



# 04 方案成效

- 使用LightGBM作为训练模型
- 静态多目标模型对月末3天进行3倍过采样，滚动单目标模型不过采样
- 设置滚动单目标模型和静态多目标模型的阈值z-score=2，融合权重为0.5。

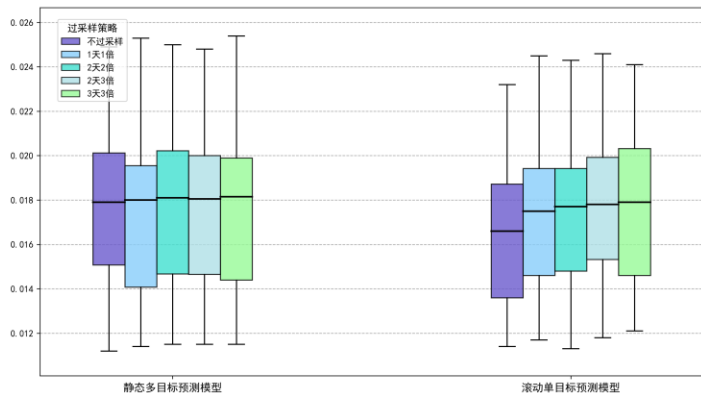
	官方_baseline	滚动单目标模型	静态多目标模型	融合模型
6月平均预测误差	0.0349	0.0167	0.01631	0.0159
7月平均预测误差	0.0380	0.0196	0.0204	0.0195

## 过采样消融实验

静态多目标模型月末2天3倍过采样效果较好，滚动单目标模型不过采样效果较好。可能是滚动模型捕捉月末大量赎回现象弱

	不过采样	1天1倍	2天2倍	2天3倍	3天3倍
静态多目标模型 平均预测误差	0.01780	0.01777	0.01776	<b>0.01766</b>	0.01785
滚动单目标模型 平均预测误差	<b>0.01732</b>	0.01741	0.01734	0.01767	0.01766

静态多目标预测模型与滚动单目标模型过采样策略对比图

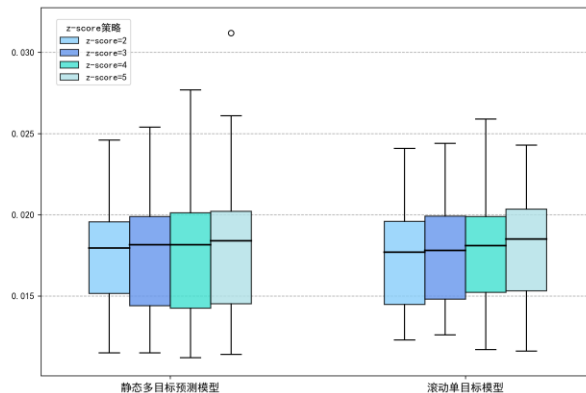


## z-score消融实验

静态多目标模型和滚动单目标模型在阈值z-score=2时效果较好，即过滤掉5%的极端数据

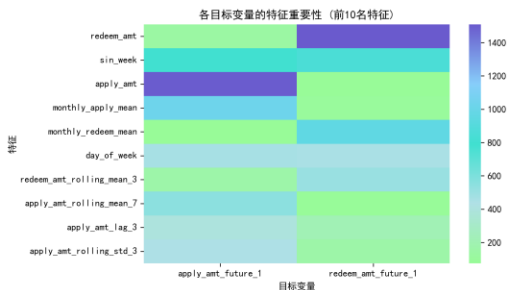
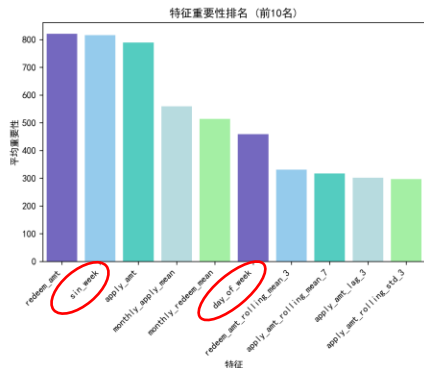
	z-score=2	z-score=3	z-score=4	z-score=5
静态多目标模型 平均预测误差	<b>0.01756</b>	0.01785	0.01797	0.01833
滚动单目标模型 平均预测误差	<b>0.01741</b>	0.01766	0.01793	0.01796

静态多目标预测模型与滚动单目标模型z-score策略对比图

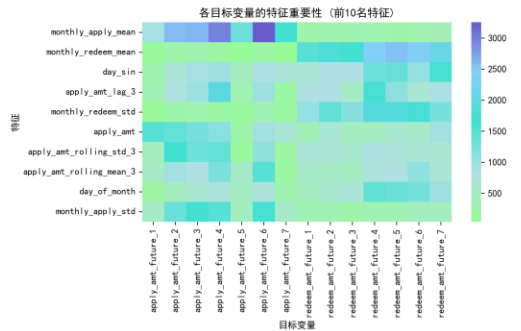
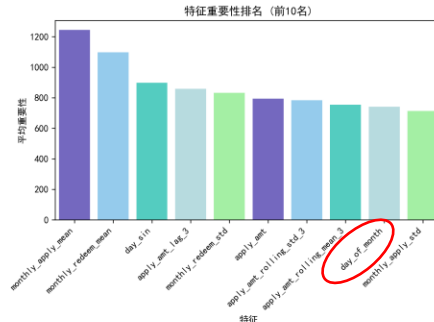


## 发现1

当月平均申购/赎回量是重要的特征。滚动单目标模型擅长捕捉基金1周内申购/赎回变化，静态多目标模型擅长捕捉基金1月内申购/赎回变化。



滚动单目标预测模型特征重要性

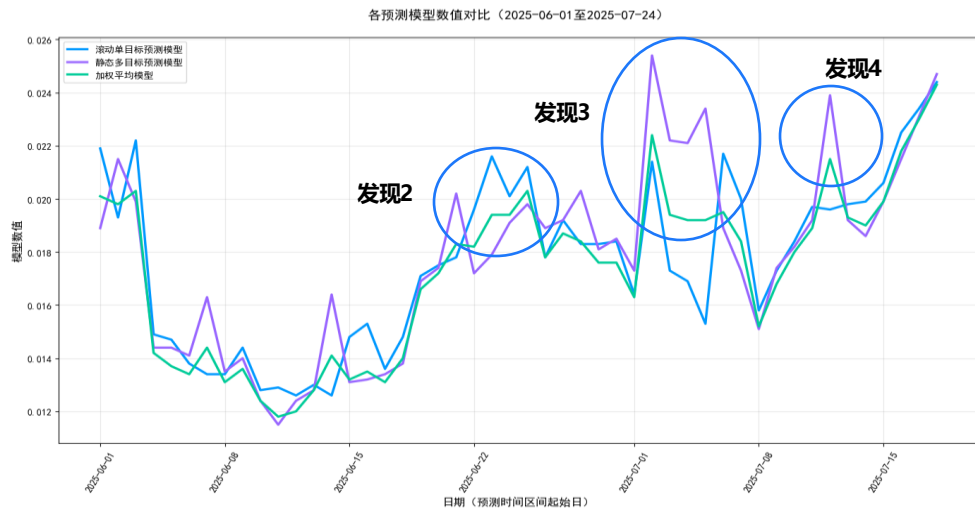


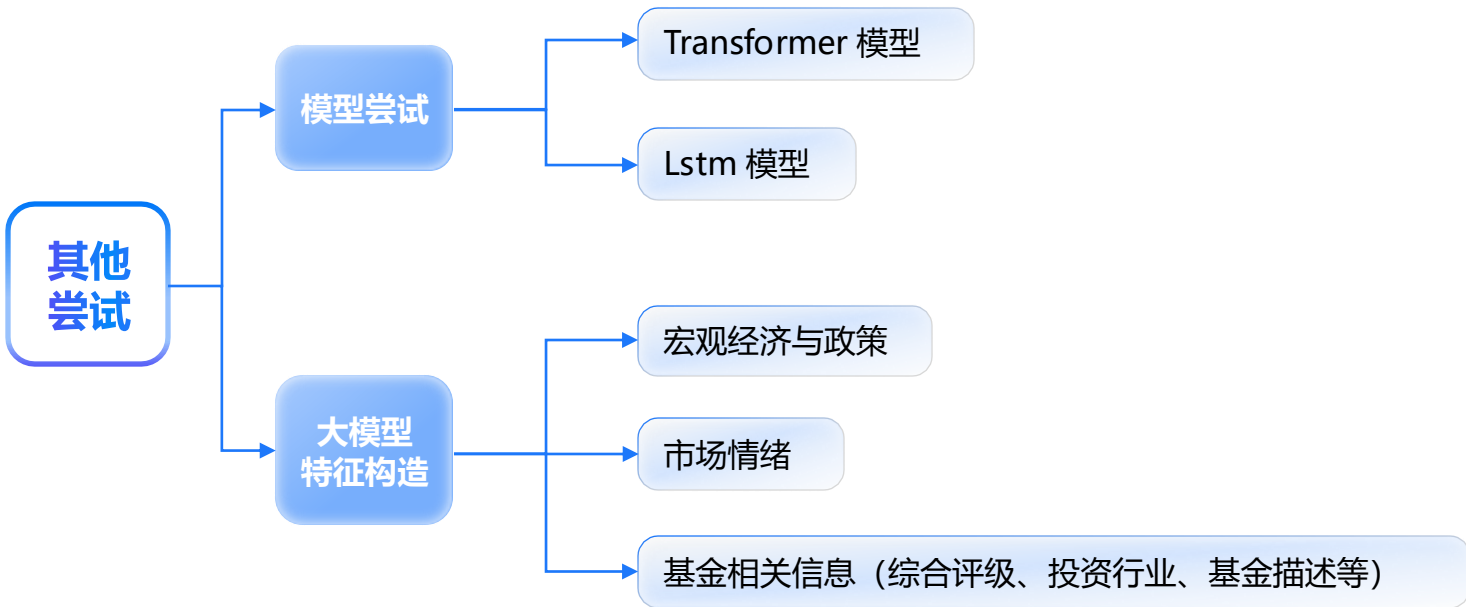
静态多目标预测模型特征重要性

**发现2** 静态多目标模型在月末表现优于滚动单目标模型，可能原因是静态多目标模型擅长捕捉基金1月内申购/赎回变化，能够学会基金在月末最后1个工作日有大量赎回的现象。

**发现3** 当预测区间属于“月末-月初”情况时，推荐使用滚动单目标模型。因为静态多目标模型的训练样本需要未来7天的申购、赎回作为标签，缺少月初的样本作为训练集，从而导致预测效果下降。

**发现4** 当基金的申购、赎回出现剧烈变化时，推荐使用滚动单目标模型。可能原因在于静态多目标模型的训练数据未包含最新1周的样本，导致模型无法捕捉到当前申购赎回的趋势变化。







# 05 技术创新

## 创新点1：大模型驱动的构建基金预测模型框架

- 省去人工设计、筛选特征的重复工作，显著缩短建模周期。同时规避主观经验偏差，提升特征质量与预测稳定性。

## 创新点2：大模型特征构造模块可拓展性强

- 可以灵活调用不同的mcp服务，从外部数据获取基金的特征。



恳请各位批评指正，谢谢！

答辩队伍：CEDC冲冲冲

## 国内首家由中央企业设立的数据产业集团

定位：数字中国建设和数字经济发展的核心战略科技力量，数字政府建设、数据安全与数据要素化治理的主力军，数据安全与数据治理技术的策源地，数据产业链的链长。

2022年12月25日揭牌 2023年4月12日正式成立

