

金融服务机构的资金流量预测

1 项目背景

某金融服务机构拥有上亿会员，并且业务场景中每天都涉及大量的资金流入和流出，面对如此庞大的用户群，资金管理压力会非常大。在既保证资金流动性风险最小，又满足日常业务运转的情况下，精准地预测资金的流入流出情况变得尤为重要。企业希望能精准预测未来每日的资金流入流出情况。对货币基金而言，资金流入意味着申购行为。资金流出意味着赎回行为。

2 项目目标

基于企业希望精确预测资金流入流出数量的需求，设定项目目标为：预测蚂蚁金服次月每天的申购总额。

3 项目步骤

3.1 工程前期准备

3.1.1 导入数据

(1) 介绍数据

用户申购赎回数据表：表中包含 2013 年 7 月 1 日至 2014 年 8 月 31 日的申购和赎回信息、以及所有的子类目信息。数据经过脱敏处理，脱敏之后的数据，基本保持了原数据分布。数据主要包括用户操作时间和操作记录，其中操作记录包括申购和赎回两个部分。金额的单位是分，即 0.01 元人民币。如果用户今日消费总量为 0，即 `consume_amt=0`，则四个子类目为空。如表 3-1 所示。

表 3-1 用户申购赎回数据表

属性	含义	示例
user_id	用户 id	1234
report_date	日期	20140407
tBalance	今日余额	109004
yBalance	昨日余额	97389
total_purchase_amt	今日总购买量 = 直接购买 + 收益	21876
direct_purchase_amt	今日直接购买量	21863
purchase_bal_amt	今日支付宝余额购买量	0
purchase_bank_amt	今日银行卡购买量	21863
total_redeem_amt	今日总赎回量 = 消费 + 转出	10261
consume_amt	今日消费总量	0
transfer_amt	今日转出总量	10261
tftobal_amt	今日转出到支付宝余额总量	0
tftocard_amt	今日转出到银行卡总量	10261
share_amt	今日收益	13
category1	今日类目 1 消费总额	0
category2	今日类目 2 消费总额	0
category3	今日类目 3 消费总额	0
category4	今日类目 4 消费总额	0

(2) 上传数据到 Python 数据挖掘建模平台

在新增数据源上，选择本地上传数据，如图 1 所示。



图 1 本地上传数据源

在本地路径上选择文件，填写在平台新建的目标表名，如图 2 所示。



图 2 本地选择文件上传

根据文件的数据，可以修改文件的字段名和类型，如图 3 所示。



图 3 字段设置

上传成功，可以在平台的数据源上查看数据，单击数据源操作的查看按钮如图 4 所示，数据预览如图 5 所示。

+ 新增数据源						请输入表名	选择状态	请选择创建时间
表名	创建人	数据来源	同步状态	创建时间	操作			
total_purchase_amt	xinyou	结构化文件	同步完成	2019-05-28 09:59:52				
credit_card	xinyou	结构化文件	同步完成	2019-05-28 09:29:59				
disodata	xinyou	结构化文件	同步完成	2019-05-28 08:46:49				
hotspotdata	xinyou	结构化文件	同步完成	2019-05-27 15:33:36				
user_dat	xinyou	结构化文件	同步完成	2019-05-27 13:59:09				

图 4 单击预览数据按钮

预览数据 (分页加载)		
names	report_date	total_purchase_a mt
1	2013-07-01	32488348
2	2013-07-02	29037390
3	2013-07-03	27270770
4	2013-07-04	18321185
5	2013-07-05	11648749
6	2013-07-06	36751272
7	2013-07-07	8962232
8	2013-07-08	57258266

共 427 条 100 条/页 < 1 2 3 4 5 > 前往 1 页

图 5 数据预览

3.1.2 新建空白工程

右击我的工程，新建一个空白的工程，如图 6 所示。

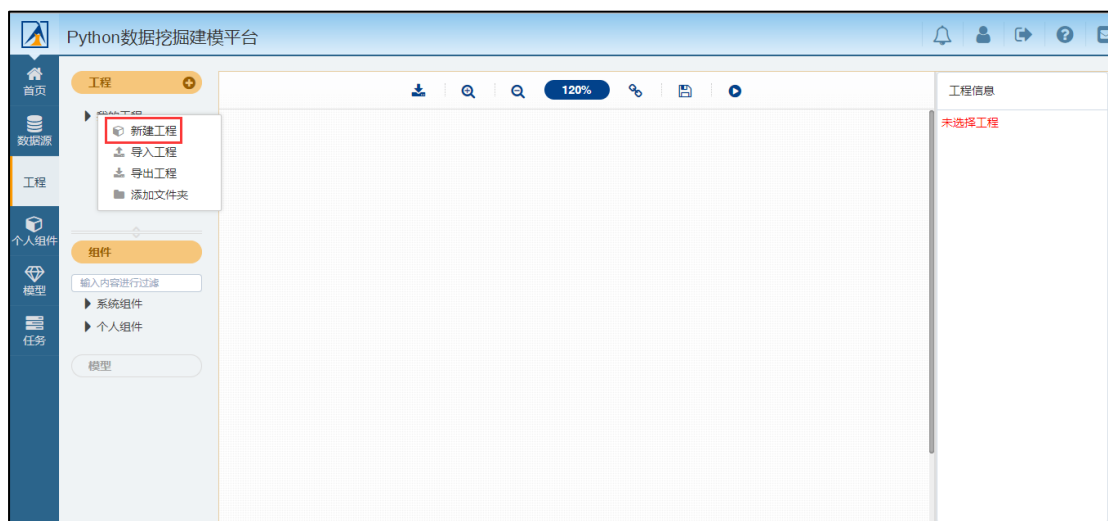


图 6 新建工程

填写工程的信息，包括工程名称和工程描述，如图 7 所示。

创建工程
✕

* 工程名称

工程描述

工程位置 ▼ 我的工程

图 7 填写工程信息

3.2 数据预处理

读取 total_purchase_amt 数据，步骤如图 8 所示。

- (1) 选择工程。
- (2) 选择输入源组件。
- (3) 拖入输入源组件。
- (4) 填写数据表名。
- (5) 单击更新按钮，更新出数据。

工程
120%

组件

系统组件

- ▼ 输入输出
- 输入源
- 输出源
- 预处理
- 统计分析
- 分类
- 回归
- 聚类
- 时序模型
- 关联规则
- 模型组件

输入源

字段属性

数据表

total_purchase_amt

字段信息

字段	类型	取值范围
names	数值	
report_date	日期	
total_purchase_amt	数值	

图 8 输入源组件

3.2.1 缺失值处理

数据可能存在缺失值，先对数据进行缺失值处理，步骤如**错误!未找到引用源。**所示。

- (1) 找到预处理→缺失值处理组件。
- (2) 拖入缺失值处理组件，并将输入源和缺失值处理组件连接。
- (3) 单击更新按钮，勾选全部数据的字段作为输出字段。
- (4) 对缺失值处理组件右键，选择运行该节点。

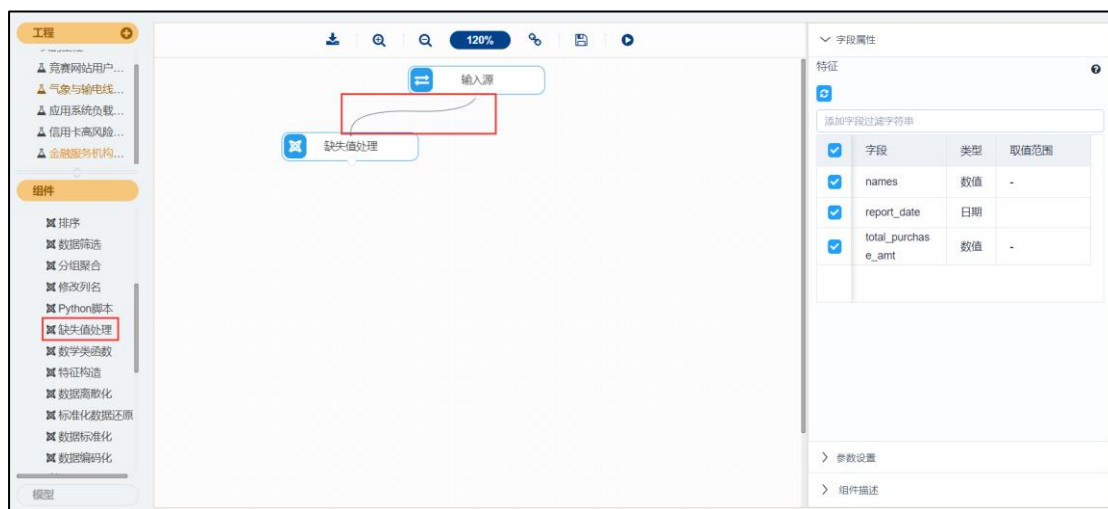


图 9 缺失值处理组件

- (5) 运行完成后，对缺失值处理组件右键，选择查看数据，如**错误!未找到引用源。**所示。

预览数据		
names	report_date	total_purchase_amt
1	2013-07-01	32488348
2	2013-07-02	29037390
3	2013-07-03	27270770
4	2013-07-04	18321185
5	2013-07-05	11648749
6	2013-07-06	36751272
7	2013-07-07	8962232
8	2013-07-08	57258266

共 427 条 25 条/页 < 1 2 3 4 5 6 ... 18 > 前往 1 页

图 10 缺失值处理结果

3.2.2 平稳性检验

选择平稳性检验，步骤如图 11 所示。

- (1) 找到统计分析→平稳性检验。
- (2) 拖入平稳性检验组件，将缺失值处理和平稳性检验组件连接。
- (3) 单击更新按钮，时序特征勾选 total_purchase_amt 字段作为检验字段。
- (4) 对平稳性检验组件右键，选择运行该节点。

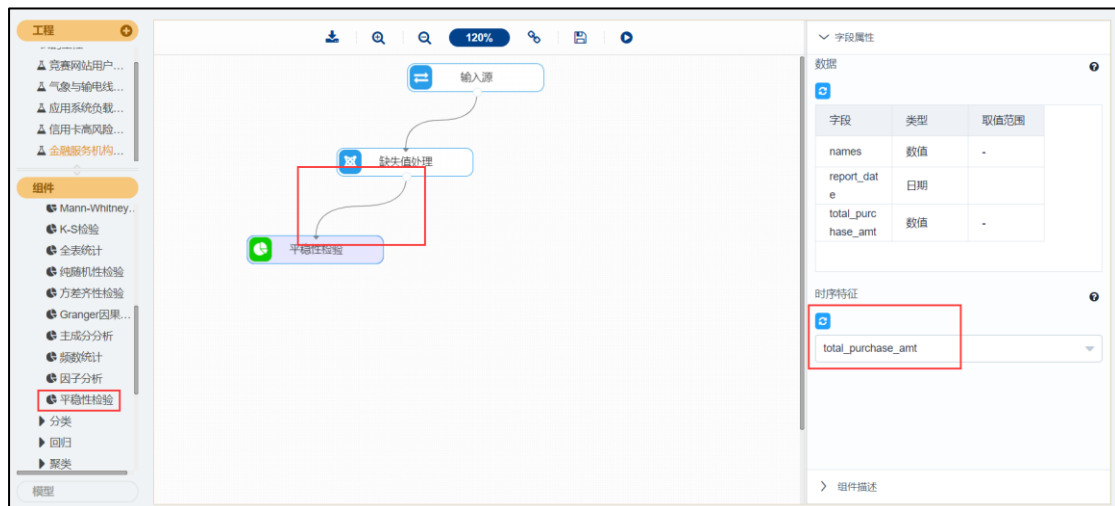


图 11 平稳性检验组件

(5) 运行完成后,对平稳性检验组件右键,选择查看报告,平稳性检验的报告结果如图 12 所示。



图 12 平稳性检验报告

3.2.3 纯随机性检验

选择平稳性检验,步骤如图 13 所示。

- (1) 找到统计分析→纯随机性检验。
- (2) 拖入平稳性检验组件,将缺失值处理和纯随机性检验组件连接。
- (3) 单击更新按钮,特征勾选 total_purchase_amt 字段作为检验字段。
- (4) 对纯随机性检验组件右键,选择运行该节点。

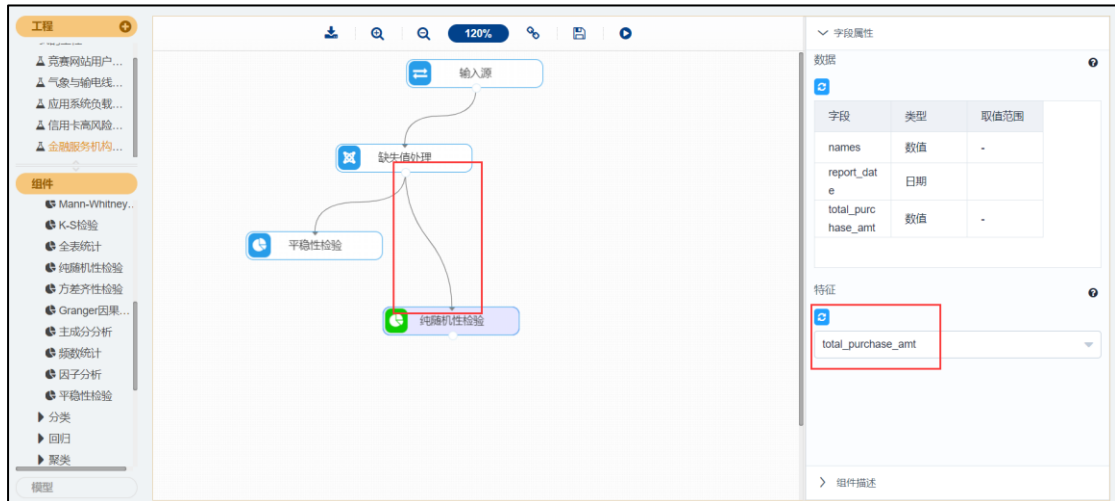


图 13 纯随机性检验组件

(5) 运行完成后，对纯随机性检验组件右键，选择查看报告，纯随机性检验的报告结果如图 14 所示。



图 14 纯随机性检验报告

3.3 模型构建

3.3.1 ARIMA 算法

选择 ARIMA 算法模型，步骤如图 15、图 16 所示。

(1) 找到时序模型→ARIMA 组件。

(2) 拖入 ARIMA 组件，将生成训练数据和 ARIMA 组件连接。

(3) 选择字段属性，单击更新数据，时序列勾选 `total_purchase_amt` 字段，时间列勾选 `report_date` 字段。

(4) 选择参数设置，设置预测周期数的值为 10，设置自回归项数 p 的值为 7，设置差分次数 d 的值为 1，设置移动平均项数 q 的值为 2。

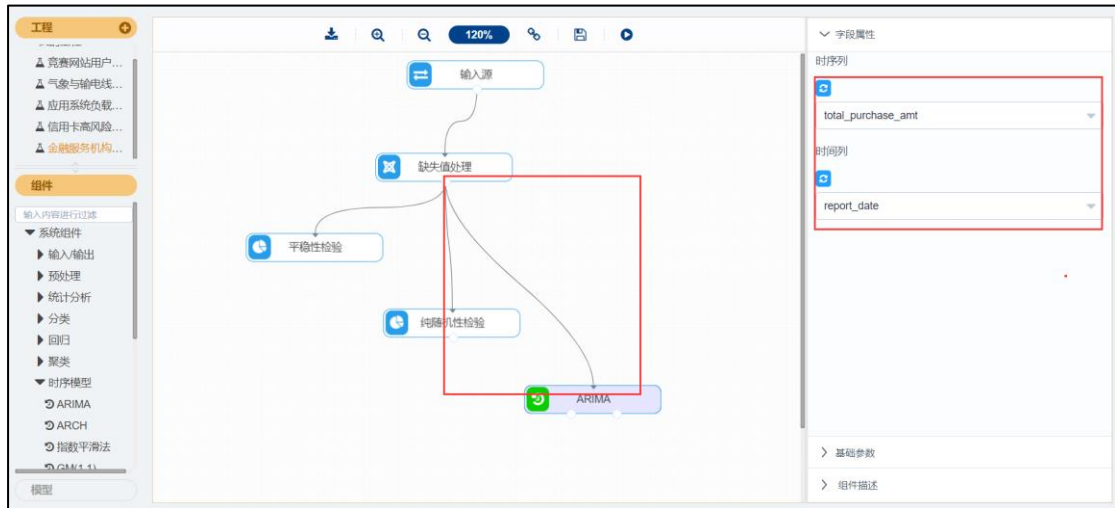


图 15 ARIMA 组件_字段属性

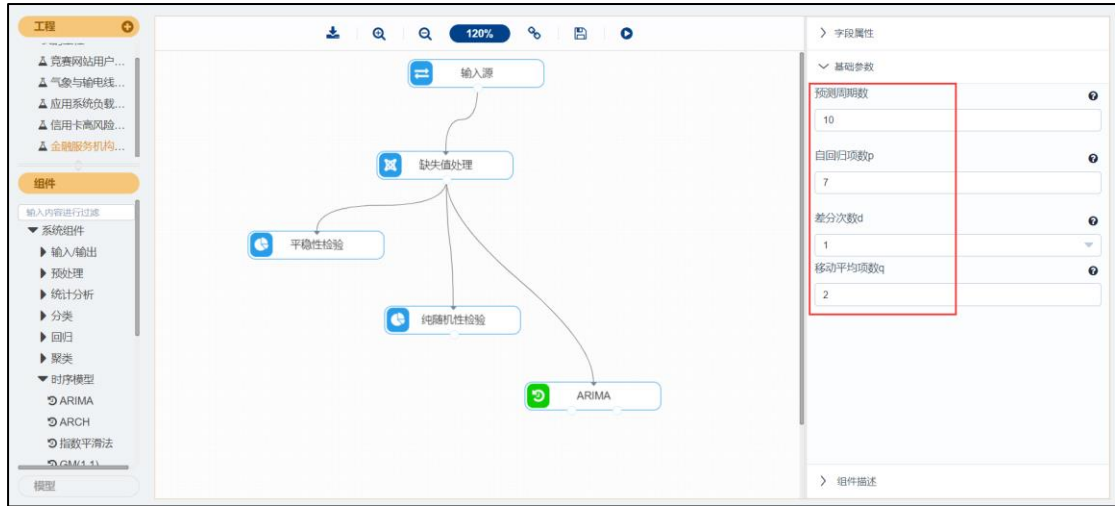


图 16 ARIMA 组件_参数设置

(5) 运行完成后，对 ARIMA 组件右键，选择查看报告，ARIMA 的报告如图 17 所示。



图 17 ARIMA 的报告