

需求：分析“银行流水数据”，识别商品信息

一、银行流水数据样例：

A										G																			
客户名称：成都望江环保科技有限公司 账号：22848401040010577 开户行：中国农业银行金堂支行																													
2	交易日期	收入金额	支出金额	账户余额	对方账号	对方户名	交易用途																						
3	20220107 13:23:49	218,959.20	0.00	769623.00	31130101040014754	重庆新思通金属材料有限公司	货款																						
4	20220108 11:17:21	92,862.00	0.00	662485.00	31161201040001686	重庆中宏市政设施有限公司	货款																						
5	20220108 12:49:42	110,937.60	0.00	973422.60	31130101040014754	重庆新思通金属材料有限公司	货款																						
6	20220110 14:25:53	211,438.20	0.00	1184860.80	31130101040014754	重庆新思通金属材料有限公司	货款																						
7	20220110 14:55:11	0.00	500,000.00	684980.80	22335201040011835	四川中再生资源开发有限公司	货款																						
8	20220110 14:55:11	0.00	15.00	684865.80			付5677账户本金500000日志号47826641手续费 (050267)																						
9	20220111 09:43:04	76,853.00	0.00	761698.80	31161201040001686	重庆中宏市政设施有限公司	货款																						
10	20220111 14:34:58	0.00	240,000.00	521698.80	6228270461235042874	陈志友	还款																						
11	20220112 20:27:02	113,534.40	0.00	336233.20	31130101040014754	重庆新思通金属材料有限公司	货款																						
12	20220113 17:41:30	0.00	300,000.00	336233.20	22335201040011835	四川中再生资源开发有限公司	货款																						
13	20220113 17:41:30	0.00	15.00	336218.20			付5677账户本金300000日志号526918517手续费 (050267)																						
14	20220117 14:17:45	0.00	100,000.00	236218.20	6228270461235042874	陈志友	还款																						
15	20220117 18:32:31	108,140.80	0.00	34339.00	31130101040014754	重庆新思通金属材料有限公司	货款																						
16	20220118 11:03:13	300,000.00	0.00	34339.00	31160201040009151	永川区供粪物资回收站	货款																						
17	20220118 12:10:16	0.00	300,000.00	34339.00	6228270461235042874	陈志友	还款																						
18	20220118 17:08:04	0.00	7,821.40	33557.60	22999901010002402	钱惠如	往来款																						
19	20220119 11:10:23	400,000.00	0.00	73557.60	0153014210012442	重庆新思通金属材料有限公司	货款																						
20	20220119 14:19:23	0.00	500,000.00	23557.60	22335201040011835	四川中再生资源开发有限公司	货款																						
21	20220119 14:19:23	0.00	15.00	23552.60			付5677账户本金500000日志号423138159手续费 (050267)																						
22	20220121 08:59:33	117,945.20	0.00	35347.80	31130101040014754	重庆新思通金属材料有限公司	货款																						
23	20220121 15:23:47	369,420.38	0.00	72288.18	31130101040014754	重庆新思通金属材料有限公司	货款																						
24	20220121 15:40:27	0.00	500,000.00	22288.18	22335201040011835	四川中再生资源开发有限公司	货款																						
25	20220121 15:40:27	0.00	15.00	22287.18			付5677账户本金500000日志号506061907手续费 (050267)																						
26	20220124 21:16:19	200,000.00	0.00	42287.18	31160201040009151	永川区供粪物资回收站	货款																						
27	20220124 21:25:27	111,837.20	0.00	534710.38	31160201040009151	永川区供粪物资回收站	货款																						
28	20220124 21:39:50	0.00	150.00	534560.38			付本账户2022年企业网银年服务费																						
29	20220126 10:25:48	0.00	400,000.00	434560.38	622846046801976972	唐世龙	还款																						
30	20220126 11:04:44	300,000.00	0.00	434560.38	31160201040009151	永川区供粪物资回收站	货款																						
31	20220126 11:51:18	0.00	15.00	434545.38	22848401940000310		还款																						
32	20220126 15:20:29	0.00	100,000.00	334545.38	6212264402011889015	唐联华	还款																						
33	20220126 15:20:29	0.00		334545.38			付5677账户本金100000日志号999999999手续费 (050273)																						
34	20220127 11:18:42	730,300.00		334545.38																									

二、模型的入参参数：

1、入参参数：

序号	参数名	说明	示例
1	res1	(1) 流水之外的数据； 包含了“bank_name”、“accout_num”、“account_name”、“idcard_num”、“agent_type”、“outside_infos” 六个字段信息； (2) “bank_name”、“accout_num”、“account_name”、“idcard_num”、“agent_type”五个字段结果为空，不用管； (3) “ outside_info” 的内容是的各个字段内容和字段的坐标信息， 其中“ doc_type” = “ ImagePDF” 的，有坐标信息；“ doc_type” = “ excel”、“ epdf” 的，不含坐标信息； 0 值是指页面中， 位置在流水之前的文本， 1 值是指位置在流水之后的文本内容；	参考下文的示例；
2	res2	(1) 流水数据, 包含"row_order"、"header_row"、“文本内容” ； (2) "row_order"代表该行数据在第几页第几行； (3) "header_row"结果无用，不用管； (4) “文本内容” 包含了该行所有单元格的位置坐标和文本内容；	参考下文的示例；
3	page_sum	(1) 代表总页数， 显示该文件一共的页数；	"007"
4	doc_type	(1) 代表文件类型, doc_type 一共分为 excel、epdf、imagePDF 三种类型； 方便算法做判断；	"imagePDF"
5	error_page	(1) 丢失的页码， 代表这几页的数据是没有提供的； 如果为空， 代表没有丢失的页码；	[]

2. 示例

1) res1 示例:

```
"res1": {
  "bank_name": "",
  "account_num": "",
  "account_name": "",
  "idcard_num": "",
  "outside_infos": [
    [
      {
        "box": [
          343.0,
          89.0,
          582.0,
          111.0
        ],
        "txt": "杭州银行账户交易明细查询"
      },
      {
        "box": [
          1.0,
          154.0,
          46.0,
          179.0
        ],
        "txt": "账号"
      }
    ],
    []
  ]
},
```

2) res2 示例:

```
"res2": [
  {
    "row_order": "001_001",
    "header_row": false,
    "1.1": "交易日期",
    "1.2": "摘要",
    "1.3": "借方金额",
    "1.4": "贷方金额",
```

```

    "1.5": "账户余额",
    "1.6": "机构号",
    "1.7": "方账老库支行 业务专用章",
    "1.8": "可 对手方名称"
  },
  {
    "row_order": "007_009",
    "header_row": false,
    "9.1": "2022-04-28",
    "9.2": "网还款",
    "9.3": "23000.00",
    "9.4": "0.00",
    "9.5": "105.37",
    "9.6": "7790",
    "9.7": "3301050260021477060",
    "9.8": "王琦晖"
  }
]

```

三. 输出参数

1、“outres1” 字段输出，包含以下内容；

该字段是当前发生账户的相关信息；

序号	字段	说明	内容示例
1	agent_type	机构类型，包含以下三种： bank（银行）， alipay（支付宝）， wechat（微信支付）；	bank
2	bank_name	如果“agent_type”是alipay\wechat的，该字段为空； 如果“agent_type”是bank，该字段需要显示； 银行机构名称参考“金融机构信息对照表”，输出内容包含编号和名称； http://doc.shargoodata.com/docs/bank20230709/bank20230709-1evq63akq38ur	GY0001 交通银行
3	accout_num	银行账号号码，支付宝账号，微信账号；	6228480329572520271
4	account_name	账号户名、支付宝账号姓名、微信账号姓名	李数
5	idcard_num	用户身份证号码，机构类型是alipay\wechat的，有这个字段，bank的没有；	330724197805240011
6	date_start	起始日期	2022/05/22
7	date_end	结束日期	2022/11/22
8	account_type	账户类型：对私、对公；	对私

2、“outres3” 字段输出，包含以下内容；

该字段是流水信息的标准固定格式输出，对表头数量和内容格式都做了统一化处理，以下是

bank、wechat、alipay 三种机构各自的输出格式；

序号	机构类型 (agent_type)	字段名	格式示例	处理说明
1	bank	交易日期	2022/03/15 (yyyy/mm/dd)	原数据如果有多个，只选择其中一个；
		收支类型	收入 (文本)	如果收入、支出两列表示的，合并成一列表示；
		交易金额	3571.42 (数值)	只显示数字，方向由“收支类型”表示；
		余额	3571.42 (数值)	
		备注 (摘要)	跨行转账 (文本)	原数据备注如果有多个，就输出多个，以备注 1、备注 2、备注 3 形式；
		对方账号	622841.....8978 (文本)	交易对方的账号；
		对方户名	张** (文本)	交易对方的户名；
2	wechat	交易日期	2022/03/15 (yyyy/mm/dd)	
		交易类型	商户消费 (文本)	
		收/支/其他	收入 (文本)	
		交易方式	中国银行 (6633) (文本)	
		交易金额	3571.42 (数值)	
		交易对方	王** (文本)	
3	alipay	收/支	其他 (文本)	除了收和支，还有“其他”；
		交易对方	天弘基金管理有限公司 (文本)	交易对方账号，其中“收支明细证明”“”样本中没有该信息，可以为空；
		商品说明	转账收款到余额宝 (文本)	“收支明细证明”“”样本中取“名称/备注”字段；
		收 / 付 款 方 式	余额宝 (文本)	当前发生账户，“收支明细证明”样本中，取自“资金渠道”字段，可以直接返回“支付宝余额”；
		金额	1.77 (数值)	交易金额，“”收支明细证明样本中取“收入”“支出”两列；并且填充“收/支”字段；
		交易日期	2022/03/16 (yyyy/mm/dd)	

3、“outres2” 字段输出，包含以下内容：

1) 该字段显示每行流水数据的分类识别结果；

序号	字段名	说明	内容示例
1	row_order	所属的页码和行序， 上半部分数字代表页码，下半部分数字代表行序	002_005
2	header_row	表头行判断， “ true” 表示该行是表头行， “ false” 表示该行不是表头行；	true
3	row_label	(1) 行分类标签，包含 “trader_nature”、 “in_spend_type”、“product_type” 三个字段； (2) “trader_nature”，交易对象的行业分类； 一共有 1382 个行业分类； (3) “in_spend_type”，收支类型，针对当前 发生账户的资金方向：收入/支出；增加显示该 字段，是因为有些流水的收支类型会显示 “其 他”，所以需要额外判断； (4) “product_type”：交易的商品分类或者叫 资金用途分类，显示交易的实物商品、虚拟商品、 服务等；一共有 38 类商品分类；	参考下面的示例；
4	account_label	(1) 行的会计分录，包含 “N-accounting_item”、 “N-accounting_entry” 、 “accounting_item”、“accounting_entry” ； (2) “N-accounting_item” 会计分录的主科 目，表示当前发生账户科目；参考《会计分录& 科目标签-会计科目》；以编号+名称的形式输出； (3) “N-accounting_entry” 代表主科目的分 录方向，用减少/增加表示； (4) “accounting_item” 会计分录的副科目， 用来表示资金用途； (5) “accounting_entry” 是副科目的分录方 向，用减少/增加表示；	参考下面的示例；

2) outres2 示例：

```
"outres2": [  
  {  
    "row_order": "001_1",  
    "header_row": "true",  
    "1.1": "收/支",  
    "1.2": "交易对方",  
    "1.3": "商品说明",  
    "1.4": "收/付款方式",
```

```

"1.5": "金额",
"1.6": "交易订单号",
"1.7": "商家订单号",
"1.8": "交易时间"
},
{
  "row_order": "001_2",
  "header_row": "false",
  "2.1": "其他",
  "2.2": "浙江网商银行股份有限公司",
  "2.3": "账户结息- 2022.11.22",
  "2.4": "",
  "2.5": "0.03",
  "2.6": "20221122102553010048250 070316309",
  "2.7": "20221122101207120700000 00000670001002013INTERE ST",
  "2.8": "2022-11-22 15:30:43",
  "row_label" :{
    "trader_nature" : " 6621 商业银行服务" ,
    "in_spend_type" : " 收入" ,
    "product_type" : A0104 利息
  }
  "account_label" :{
    "N-accounting_item" : "Z07 账户余额" ,
    "N-accounting_item" : " 增加" ,
    "accounting_item": "S02 营业外收入",
    "accounting_entry": "增加"
  }
}
],

```

4、所有输出参数

1) 所有输出参数

序号	参数名	说明	输出内容
1	success	是否正常执行成功; true: 代表执行成功 false: 代表执行失败	true
2	error_code	如果 success=false 时的返回错误码 401: 文件格式不符合要求 402: 超过文件大小限制 -1: 未知错误 工程师添加其他错误码	401
3	message	报错信息, 与 error-code 对应	文件格式不符合要求;

4	data	识别数据, 包含 page_sum (沿用输入字段内容), doc_type (沿用输入字段内容), outres1, outres3 outres2,	参考返回示例
---	------	---	--------

2) 返回参数示例

```
{
  "success": "true",
  "error_code": "",
  "message": "",
  "data": {
    "page_sum": 083,
    "doc_type": "epdf",
    "outres1": {
      "agent_type": "alipay",
      "account_num": "2504767094@qq.com",
      "account_name": "王克",
      "idcard_num": "370213199907186016",
      "account_type": "对私",
      "begin_date": "2022/05/22",
      "end_date": "2022/11/22"
    },
    "outres2": [
      {
        "row_order": "001_1",
        "header_row": "true",
        "1.1": "收/支",
        "1.2": "交易对方",
        "1.3": "商品说明",
        "1.4": "收/付款方式",
        "1.5": "金额",
        "1.6": "交易订单号",
        "1.7": "商家订单号",
        "1.8": "交易时间"
      },
      {
        "row_order": "001_2",
        "header_row": "false",
        "2.1": "其他",
        "2.2": "浙江网商银行股份有限公司",
        "2.3": "账户结息- 2022.11.22",
        "2.4": ""
      }
    ]
  }
}
```

```

"2.5": "0.03",
"2.6": "20221122102553010048250 070316309",
"2.7": "20221122101207120700000 00000670001002013INTERE ST",
"2.8": "2022-11-22 15:30:43",
  "row_label" :{
    "trader_nature" : " 6621 商业银行服务" ,
    "in_spend_type" : " 收入" ,
    "product_type" : A0104 利息
  "account_label" :{
    "N-accounting_item" : " Z07 账户余额" ,
    "N-accounting_item" : " 增加" ,
    "accounting_item": "S02 营业外收入",
    "accouting_entry": "增加"
  }
}
],
"outres3": [
  {
    "row_order": "001_1",
    "header_row": "true" ,
    "1.1": "收/支",
    "1.2": "交易对方",
    "1.3": "商品说明",
    "1.4": "收/付款方式",
    "1.5": "金额",
    "1.6": "交易日期"
  },
  {
    "row_order": "001_2",
    "header_row": "false" ,
    "2.1": "其他",
    "2.2": "浙江网商银行股份有限公司",
    "2.3": "账户结息-2022.11.22",
    "2.4": "",
    "2.5": 0.03,
    "2.6": "2022/11/22"
  }
]
}

```

四、交付内容

- 1、源代码（包含 dockfile）
- 2、镜像文件

3、部署文档（包含分别采用 GPU 或 CPU 启动方式的说明）

4、接口文档