

周工作总结0408-0412——李东升

本周完成了房评审核的自测case程序开发，内部指标线上化，还有不当内容的几个规则，以及贷款部分的规则开发，机审覆盖率从8%提升至24%，业务方表示了赞赏。同时看点这边完成了不过审文章分析，发现问题所在，并制定了本次迭代下周的排期。

技术积累

1. 学会用postman发http请求测试服务。

2. hive建表报权限错误：

hive的建表语句，跟mysql基本差不多，但是遇到一个权限问题，FAILED: Execution Error, return code 1 from org.apache.hadoop.hive.ql.exec.DDLTask.
MetaException(message:Got exception:
org.apache.hadoop.security.AccessControlException Permission denied:
user=aisearch, access=WRITE,
inode="/user/hive/warehouse/ssc_audit_house_comment_algorithm_iters_da":b
igdata:supergroup:drwxrwx--x

在强哥的指点下，才发现要在表名前加上database的名字，因为里面涉及用户的权限问题，我在101work账户上操作，使用的用户是aisearch，而这个用户只能在aisearch库里进行操作，所以最终的语句是：

```
CREATE TABLE IF NOT EXISTS
`aisearch.ssc_audit_house_comment_algorithm_iters_da`(
  `id` bigint COMMENT '自增序号',
  `dataset_v` varchar(64) COMMENT '以时间格式标记的数据集版本, 比如
20190405',
  `model_type` varchar(64) COMMENT '模型类型, 以下划线拼接, 比如
rule_touzi表示单类规则投自己, rule_svm表示规则加svm',
  `thresh` float COMMENT '模型阈值',
  `precision` float COMMENT '方法的准确率',
  `recall` float COMMENT '方法的召回率',
  `coverage` int COMMENT '方法的覆盖数',
  `pt` date COMMENT '记录测试时间, 时间相同且数据版本号相同的为同一组实验')
ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t'
TBLPROPERTIES ('creator'='lidongsheng004', 'crate_time'='2019-04-
08')
;
```

3. git追踪远程分支

先用git branch -a 查看远程分支, 然后在本地git branch branchname建立一个新的分支, 最后用git branch --set-upstream-to=origin/ branchname命令来绑定到远程分支上。

4. json_tools对比字典或者json文本

json_tools是一个对比工具, 但是有一个问题要注意, json_tools.diff(arr1, arr2)中的两个参数的类型不同会导致结果不对, 比如一个是字典型, 一个是字符串型, 返回的就是两个变量完全不同。操作中, 字符串型比字典型多两个引号。

5. python中字符串转成dict可以用eval函数直接转。

6. 学了下用configparser,还有一个就是把配置文件写到模板里, tpl格式的文件里, 然后用sed命令替换模板里的一些占位符, 换成需要用的参数。

方案反思

1. 指标线上化

本周在做数据分析的过程中, 发现丢失数据了, 在3月19号之后就拿不到某些原因的不过审文章了, 经过跟思南确认, 原来是他们改了code, 并没有跟我们同步。所以说指标一定要线上化, 随时能看到变化才能更好地指导工作, 像这种对方连口径改了我們都不知道, 是很影响效率的。

2. 日志回收也可以加报警

之前我们都是完全依赖grafana系统管理报警。但是本周我从日志回收邮件中发现了一个问题，有103篇文章没有绑定上beike_id，经过确认，是凤凰网的链接换了，跟我们库里的名字不一样了，所以没绑定上。这个问题明明可以在脚本统计好以后就发企微报警的，为什么要等到邮件回收了，人再去看呢，这个多亏于腾提醒了我，我在周四把这个功能已经加上，并开始测试阶段了。

个人反思

本周重新定了OKR，现在手里就是三个工作了，看点和房评审核和房源问答。之前不明白OKR导向的工作，现在知道要分析工作完成的情况，跟OKR比还差多少，最大的改进方向在哪里，事情的优先级是什么，相信明白了这些，之后的工作会更高效一些。