**01 /评分**

参赛者提交的结果文件中包含对所有用户购买意向的预测结果。对每一个用户的预测结果包括两方面： （1）该用户2016-04-16到2016-04-20是否下单P中的商品，提交的结果文件中仅包含预测为下单的用户，预测为未下单的用户，无须在结果中出现。若预测正确，则评测算法中置label=1，不正确label=0； 选手从数据中自行组成特征和数据格式，自由组合训练测试数据比例。 （2）如果下单，下单的sku\_id （只需提交一个sku\_id），若sku\_id预测正确，则评测算法中置pred=1，不正确pred=0。 对于参赛者提交的结果文件，按如下公式计算得分： upfile

此处的F1值定义为： upfile

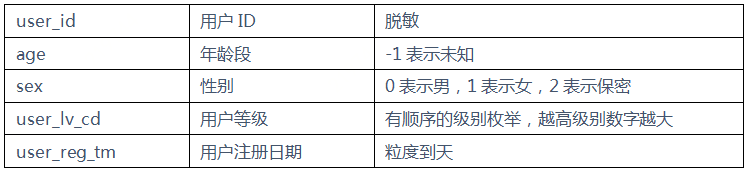
upfile

其中： Precise为准确率，Recall为召回率； F11是label=1或0的F1值，F12是pred=1或0的F1值。

**02 /数据描述**

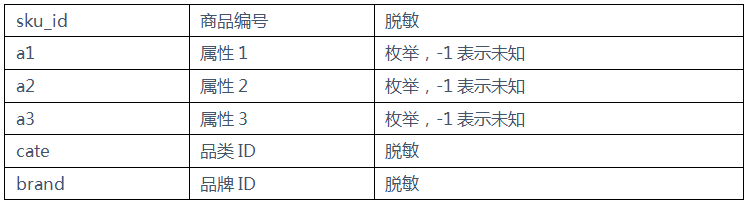
* 训练数据部分： 提供2016-02-01到2016-04-15日用户集合U中的用户，对商品集合S中部分商品的行为、评价、用户数据； 提供部分候选商品的数据P。 选手从数据中自行组成特征和数据格式，自由组合训练测试数据比例。
* 预测数据部分： 2016-04-16到2016-04-20用户是否下单P中的商品，每个用户只会下单一个商品；抽取部分下单用户数据，A榜使用50%的测试数据来计算分数；B榜使用另外50%的数据计算分数(计算准确率时剔除用户提交结果中user\_Id与A榜的交集部分)。 为保护用户的隐私和数据安全，所有数据均已进行了采样和脱敏。 数据中部分列存在空值或NULL，请参赛者自行处理。
* 符号定义： S：提供的商品全集； P：候选的商品子集（JData\_Product.csv），P是S的子集； U：用户集合； A：用户对S的行为数据集合； C：S的评价数据。
* 数据模型

（1）用户数据



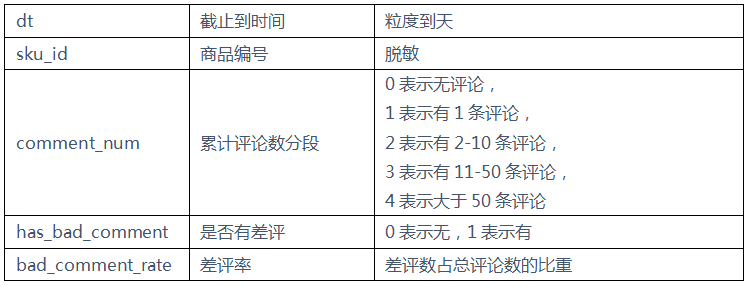
upfile

（2）商品数据



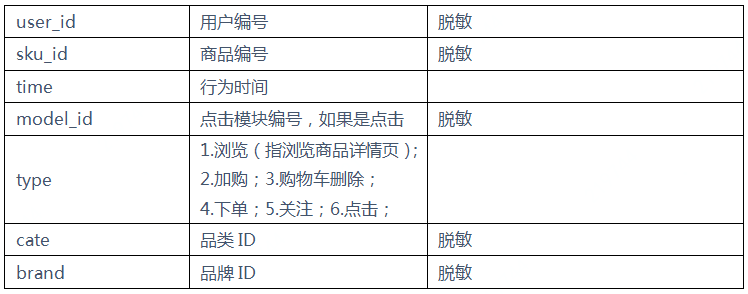
upfile

（3）评价数据



upfile

（4）行为数据



upfile

**03 /任务描述**

参赛者需要使用京东多个品类下商品的历史销售数据，构建算法模型，预测用户在未来5天内，对某个目标品类下商品的购买意向。对于训练集中出现的每一个用户，参赛者的模型需要预测该用户在未来5天内是否购买目标品类下的商品以及所购买商品的SKU\_ID。评测算法将针对参赛者提交的预测结果，计算加权得分。

**04 /提交作品要求**

初赛提交CSV结果文件，进入复赛时提交源代码。 初赛提交CSV文件中包含对有购买意向的用户所购买商品的预测结果，字段如下： user\_id：用户ID，保证唯一，请勿在一次提交的结果文件中包含重复的user\_id sku\_id：商品集合P中的商品ID，请勿在同一行中提交多个sku\_id 对于预测出没有购买意向的用户，在提交的CSV文件中不要包含该用户的信息。 提交结果demo截图如下： 