**融360赛题——多金融场景下的模型训练**

**赛题背景分析：**融360是中国最大的网络贷款平台，平台的一端是亿级别有借款需求的小微企业和个人消费者，另一端是有贷款资金的万级别的金融机构（银行、小贷、担保、典当等）和百万级的金融产品。通常，用户进入我们的平台后，自行选择最合适的金融机构和金融产品进行申请。因此，融360平台积累了大量不同时期、不同金额、不同期限、不同利率、不同市场环境下的借贷订单。在风控建模过程中，如何选择最合适的样本针对当前市场环境下特定金融产品建模，关系到所建模型在实际业务中的预测能力，直接影响金融机构的盈利能力。为此，我们提供平台上的部分数据，希望选手根据数据，解决实际问题。

**具体内容：**

**训练样本：**包括从2017.4.1到2018.5.1不同金额、不同期限、不同利率的金融产品样本，并给出每个样本的类型（属于大额分期贷或小额现金贷产品）是否逾期。（约10万样本）

**验证样本：**2018.1.1到2018.5.1机构A的产品，验证集不提供样本是否逾期，参赛选手自行完成是否逾期预测后，可以提交至比赛平台评估结果。（约2万样本）

**测试样本：**与验证样本来源相同且同分布。测试集不提供样本是否逾期，参赛选手只能在比赛最后的评比阶段将预测结果提交至比赛平台评估，且只能提交一次。（约2万样本）

**数据分析：**

由于官方给出的数据特征是6810个特征，并且这些特征并没有解释和含义，所以这就在造成了在开发特征上的困难；另外从特征的组成上来看6810个特征已经比较多了，所以该赛题主要目的是在场景建模上。

**特征工程：**因为训练集和测试集的columns的排列不一致，所以先将train和test数据集的columns以同样的方式排列。另外从时间列提取month、day、weekday。

**建立模型：**在建模上采用不同的参数、不同的接口、不同的产品建立不同的模型，经过线上反馈结果表明一下五种模型的结果是最优的，最终将这五种模型通过两两融合的方式，得到最终的结果。其五种模型的参数如程序中所示。