***2017年11月 – 至今*** **北京快快网络技术有限公司（真融宝） 普惠金融事业部 数据组算法工程师**

* 参与财小仙平台的小额现金贷项目，负责首贷模型的部分数据分析（评估模型的区分能力，分析单个特征和联合特征对目标的区分度，计算特征之间的相关性，通过数据联合建模评估第三方数据带来的增益）。通过案例分析协助机器拒绝规则的探索。
* 参与大额分期现金贷贷项目的探索，分析有较大贷款额度的用户特征；通过建立评分卡模型筛选优质用户进行放贷；通过与同盾征信联合建模筛选出适合做大额放贷模型的变量。
* 负责小贷项目中续贷订单的评分卡模型的开发（包括实现评分卡，回溯新颜数据接口，特征选择）

1. 每天订单量：之前一个月放几亿甚至上十忆元，现在缩减至千万级别元
2. 有哪些第三方数据：新颜 天机 同盾 淘金云等
3. 关于哪些方面的数据：信用分数，本平台历史还款行为，其他平台的借款行为，运营商数据（催收电话），学历工作 银联消费行为 社保公积金 短信 手机app 关联信息，银行的数据，
4. 模型评估指标：ks auc psi（样本或模型稳定指数） 精确率 召回率 准确率 混淆矩阵 包外估计
5. 模型目标是：是否逾期，逾期天数3天内，15天或者7天外
6. 最终目标是：减少逾期率，减少坏账率，提高机器自动审核比率
7. 首贷用到的特征有哪些：
8. 特征选择方法：iv 随机森林模型重要性删除 xgboost筛选 pearson相关系数
9. 有哪些拒绝规则：芝麻分 工作 最大逾期天数段超过30天 近期是否有催收电话等，评估规则：误伤
10. 两个方面评估第三方数据：计算iv值和建模分析看auc增长与否
11. 较大额度的用户特征分析：案例分析法，计算IV值
12. 评分卡模型用到的特征有哪些
13. 联合建模的过程，最终选的变量有哪些
14. 续贷模型的变量有哪些，专家卡？
15. 相关模型介绍：

* Lr（评分卡）：跨时段验证，测试集和训练集划分比例，不平衡问题？调参过程？数值算法
* 线性回归
* 决策树分类和回归：
* Bagging之RF：包外估计（自助采样法）随机森林的经验参数范围调参过程（树的棵数100，树最大深度20和分裂节点样本最少数，分裂节点样本最少数和叶子节点样本最小数，最大特征数） 特征重要性（袋外数据误差的平均值）
* Boosting之Adaboost：
* 提升树：
* Gbdt（梯度提升树）：
* Xgboost：筛特征；http://blog.csdn.net/sb19931201/article/details/52557382
* Ensemble之Stacking：
* NB
* SVM
* kNN
* 优化算法：GD SGD Adam AdaGrad
* PCA HMM CRF EM