

尊敬的潘蕊老师：

您好！

我们是一支想要参加本年度大学生创新创业比赛的统计与数学学院的学生小组，根据所学知识以及对我国当代社会重点问题的分析，我们组经讨论初步确定以个人信用评级和 P2P 网贷为切入点，充分发挥所学的统计知识，运用统计软件来探究更加适合于 P2P 网贷的个人信用评级模型。

之所以想要找您作指导教师是因为这个课题与现实关联性较强且涉及诸多数据处理技术，而我有幸在狗熊会人才计划中接受过您的教导，深感您是一位认真负责、能力极强，并对统计学的现实应用十分看重的教师，相信有您的指导我们的研究课题会逐步完善，更重要的是我们也可以在研究过程中从您身上不断学习，逐步了解科研是如何做的，通过这次比赛做到真正有所收获，我们相信这才是更加有意义的。

为了节省您的宝贵时间，在 **下页 PDF 中附上了本组基本研究思路的简短版本**。受限于篇幅及个人水平，有些表述可能不是很清晰，全体组员期待更加系统而又深度的线下交流。

最后，万分期待能够得到您的指导！

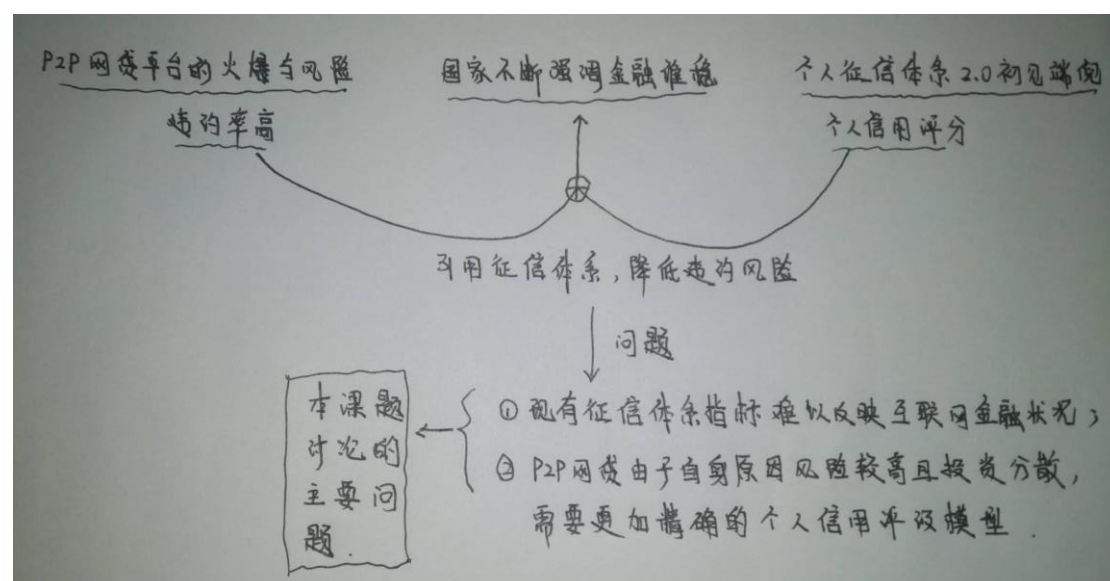
小组成员：

经济统计学 16 郝建锋、李小高、王燕坪、施霁珂

统计学 16 陈步

课题名称：P2P 网贷中个人信用评级模型的构建——基于 voting 方法的多模型融合

研究思路：



◆ 申报阶段

首先结合对传统金融机构的个人征信体系和互联网金融的背景的分析,指出传统的个人信用评分体系并不适用于如今的互联网金融征信体系,需要加以修正和优化,引出对指标体系的优化¹,核心是将平台信息(投标人数,违约历史,认证等级,借款规模,还款利率,逾期次数等)引入现有指标体系;

然后通过对 Kaggle 网站上较为规范传统的 Give me some Credit 数据集²分别采用单一模型(logit 模型、随机森林模型、Adaboost 模型)和基于 voting 方法的融合模型得出分类预测结果,并利用 ROC 曲线和 AUC 的评价体系进行比较,得出融合模型的预测精度更优的结论。

结合 P2P 网贷自身的高风险性以及最近国家不断强调的互联网金融稳中求进的方针来看,融合模型由于预测精度更高,可以更加有效的控制风险,是更符合 P2P 网贷背景下的信用评级模型。

◆ 实践阶段

1 参考文献:

[1]王会娟,廖理.中国 P2P 网络借贷平台信用认证机制研究——来自“人人贷”的经验证据[J].中国工业经济,2014(04):136-147.

[2]顾慧莹,姚铮.P2P 网络借贷平台中借款人违约风险影响因素研究——以 WDW 为例[J].上海经济研究,2015(11):37-46.

2 <https://www.kaggle.com/c/GiveMeSomeCredit/data>

根据前面建立的指标体系对人人贷网站的散标数据³利用爬虫手段进行收集（已查看 robots 协议⁴，该网站散标数据允许爬取），进一步利用该数据验证融合模型相较单一模型的优越性。

预想成果：融合模型相较而言是更符合 P2P 网贷背景下的信用评级模型，在贷款申请阶段利用融合模型进行判断，能更加有效降低平台违约率，降低金融风险；同时对变量重要性进行分析，为平台收集用户信息提供方向引导，也为借款人提升自身信用度提供了途径。

3 <https://www.renrendai.com/pc/loan.html>

4 <https://www.renrendai.com/robots.txt>