# 计算机科学与技术2015级专业实训

**指导老师：白璐 联系方式:** [**bailucs@cufe.edu.cn**](mailto:bailucs@cufe.edu.cn)

# 题目：基于随机森林算法的P2P网络借贷借款人违约预测系统

# 1. 基本介绍

P2P网络借贷是基于互联网平台的个人直接借贷，作为互联网金融典型应用模式，自2005年全球首家P2P网络借贷平台Zopa在伦敦上线，2007年中国第一家P2P平台拍拍贷成立，P2P网络借贷在全球范围内蓬勃发展，其在解决个人及中小企业融资难问题上取得了广泛的应用前景。然而，伴随着业务量的发展，平台倒闭、跑路、借款人违约等现象亦层出不穷。其中，借款人的信用风险是P2P网络借贷运营过程中面临的主要风险之一。提出一套先进的针对P2P网络借贷信用风险的管理技术对促进行业健康有序发展具有重要的现实意义。

本案例针对P2P网络借贷行业进行分析，针对平台现有借款人和新进入的借款人的不同特点，提出相应的信用风险评估模型。

# 2. 开发坏境（推荐）：

1. Python
2. Java
3. JSP技术

# 3. 交付物：

* 可以运行和演示的系统
* 源代码
* 分析和设计文档，包括

需求文档：用例图和用例、补充的说明等

分析文档：分析模型，例如系统顺序图、领域模型，SSD，操作契约等

设计文档：设计模型，决策树算法说明文档，算法伪码，相关算法设计使用的理由，参考文献脉络及与本工作的关系。例如，使用决策树模型进行违约率预测，应说明为何使用决策树算法、算法原理及特点、使用决策树算法进行违约率预测的步骤、算法伪码等。

# 4. 功能描述：

**4.1 题目：基于随机森林算法的P2P网络借贷借款人违约预测系统**

* 设计数据管理层，管理P2P网络借贷平台的实际交易数据，可利用现成的DBMS(如MySQL，SQLserver等)，也可利用其它的数据管理技术；
* 实现足球转会市场多粒度预测，即完成宏观市场及微观市场的交易额度、频度及交易模式；
* 实现任意选择两个球队间交易额度、频度及交易模式；

**具体的功能模块划分如下：**

1. **数据获取及管理层设计**

* 采取爬虫算法获取来自人人贷的实际交易数据(时间范畴为过去1年)
* 将人人贷数据和Prosper数据导入数据库，设计CRUD(create, retrieve, update, delete)操作支持上层应用。 并在此基础上做好数据的预处理。

**2) 借款人违约预测**

* 已有借款人的借款违约率预测
* 新进借款人的借款违约率预测

[注]：新进借款人的信用风险评估，只选取申请借款数为1，且无还款表现记录的借款标。

**3) 结果交付及可视化**

* 展示借款人相关信息，并展示系统给出的违约率预测结果等。

## 5. 额外要求

1. **使用5-fold validation数据来选择训练数据和测试数据**；
2. **预测准确率采用AUC及F1作为衡量标准**；
3. 3个同学一个综合项目组，每人负责自己的独立模块，选举一个综合组长协调系统整合，具体的人员组织自己商量；
4. 每个同学做一个小题目，能独立演示演示自己功能，又需要与综合组长协调，完成完整功能；
5. 本题目不限技术，以实现需求为准。