

第七届数学中国数学建模网络挑战赛

地址：数学中国数学建模网络挑战赛组委会
电话：0471-4969085

邮编：010021

网址：www.tzmcm.cn
Email：2014@tzmcm.cn

第七届“认证杯”数学中国

数学建模网络挑战赛 承 诺 书

我们仔细阅读了第七届“认证杯”数学中国数学建模网络挑战赛的竞赛规则。

我们完全明白，在竞赛开始后参赛队员不能以任何方式（包括电话、电子邮件、网上咨询等）与队外的任何人（包括指导教师）研究、讨论与赛题有关的问题。

我们知道，抄袭别人的成果是违反竞赛规则的，如果引用别人的成果或其他公开的资料（包括网上查到的资料），必须按照规定的参考文献的表述方式在正文引用处和参考文献中明确列出。

我们郑重承诺，严格遵守竞赛规则，以保证竞赛的公正、公平性。如有违反竞赛规则的行为，我们接受相应处理结果。

我们允许数学中国网站(www.madio.net)公布论文，以供网友之间学习交流，数学中国网站以非商业目的的论文交流不需要提前取得我们的同意。

我们的参赛队号为：2351

参赛队员（签名）：

队员 1：呼和

队员 2：张雅丽

队员 3：王政

参赛队教练员（签名）： 马壮

参赛队伍组别：本科

第七届数学中国数学建模网络挑战赛

地址：数学中国数学建模网络挑战赛组委会
电话：0471-4969085

邮编：010021

网址：www.tzmcm.cn
Email：2014@tzmcm.cn

第七届“认证杯”数学中国

数学建模网络挑战赛 编号专用页

参赛队伍的参赛队号：（请各个参赛队提前填写好）：

2351

竞赛统一编号（由竞赛组委会送至评委团前编号）：

竞赛评阅编号（由竞赛评委团评阅前进行编号）：

第七届数学中国数学建模网络挑战赛

地址：数学中国数学建模网络挑战赛组委会
电话：0471-4969085

邮编：010021

网址：www.tzmcm.cn
Email：2014@tzmcm.cn

2014 年第七届“认证杯”数学中国 数学建模网络挑战赛第二阶段论文

题 目 土地储备方案的风险评估

关 键 词 决策树, C4.5, 土地收储, K—均值聚类, 因子分析,

多项式拟合, 蒙特卡洛模拟, 拟合优度检验, 置信区间检验, 打分模型

摘 要:

本研究应用决策树 C4.5 算法结合 k—平均聚类算法对经过数据挖掘预处理后的土地储备贷款挖掘数据集中的数据建立打分模型得到 8 个高风险样本, 序号分别为 47、17、22、35、18、65、6、7。我们根据附件三中 5 个部门专家打分结果我们进行修正模型, 首先, 对五个部门的专家打分分别和所有 36 个属性列做相关性检验, 得到的相关性系数显示的相关性比较低, 我们通过 SPSS 对五组专家打分做了因子分析得到了 4 个主要因子 F_1, F_2, F_3, F_4 的累积贡献为 89.938%, 又通过 MATLAB 求得五组专家打分相关系数矩阵的特征向量, 基于四组主成分我进行运算得到综合因素模型, 简称综合打分模型。通过 K-S 正态性检验对综合打分模型给出的打分结果进行检验, 得到其满足正态分布的特点。同样得到了 8 个高风险样本, 序号分别为 47、17、23、49、18、39、10、5。接着, 我们根据综合评分的结果, 找出评分较低的部分项目查看其在打分模型之前定义过的不同财务内部收益率下出现高风险样本个数的分布情况, 得出所有情况对应的风险权重为 0.083、0.143、0.1、0。对打分模型修正后的结果得到 8 个高风险样本, 序号分别为 47、17、23、49、18、39、6、7, 已经很接近综合打分模型结果了并且通过了正态分布检验, 新的打分模型的所有项目打分成绩与综合模型给出的分数相关系数为 85.26%, 对于社会企业机关这个系数可以说明我们的打分模型可以有效刻画各部门专家打分的结果, 并对未来新的样本打分。对于附件二中土地收储面积、FNPV、当年现金流出、预期收益这四个指标列会随时间变动, 而且他们近几年的变动趋势是逐年上升, 已经通过多项式拟合和线性拟合进行了刻画, 并给出了拟合优度。进一步修改了决策树模型, 使得它可以在这些指标变动时给出准确评估, 并给出了这四项指标的变化区间。我们的打分模型具有较高的实际应用前景和价值。

参赛队号： 2351

所选题目： C 题

参赛密码 _____
(由组委会填写)

第七届数学中国数学建模网络挑战赛

地址：数学中国数学建模网络挑战赛组委会
电话：0471-4969085

邮编：010021

网址：www.tzmcm.cn
Email: 2014@tzmcm.cn

英文摘要

Our explore use the pre-processing data to combine the decision tree C4.5 algorithm and the K-means clustering method, to build the scoring model for the risk assessment of land store, and we made conclusion that top 8 highest risk are:47、67、22、35、18、65、6、7.we modify scoring result according to attachment three, first of all, we use each department scoring result and 36 cols of the attachment two to make correlation analysis, we conclude that attachment three are compositive compute result, and we analysis five department use the factor analysis method. We conclude that the cumulation contributions ratio is 89.938%, and gives the comprehensive scoring result and the highest risk top 8:47,17,23,49,18,39,10,5. According to this result, we conclude that the different risk level's weight are 0.083 、 0.143、 0.1、 0, so we calculate new score by the weight and conclude that the highest risk top 8 :47,17,23,49,18,39,6,7. It's very significant and the comprehensive compute result and the new result's correlation is high. According to analysis result of attachment two, we found that the area of land store, FNPV, cash flows and expected gains have high correlation with the year of this project, and data of each col is rise every year. so we modify the scoring model and give more reasonable score. Our model is significant and have great apply value.