# 金融贷款中企业多指标信用评估案例教学指导手册

**教学适用的课程：**《大数据架构与技术》

**教学适用的专业：**电子信息

**教学目的与用途：**学习实际金融风控场景下的大数据应用；掌握大数据预处理、数据分析相关知识点；掌握大数据处理与分析的基本流程。

**教学内容：**

(1) 理论依据：本案例基于大数据技术进行金融贷款中企业多指标信用评估。案例整合了多种的数据挖掘与机器学习方法，进行数据预处理与基于决策树与随机森林算法的企业多指标信用评估。

(2) 涉及知识点：数据预处理、数据可视化、随机森林分类、MapReduce。

(3) 分析路径：首先部署Hadoop+Spark大数据处理平台；然后对企业数据进行预处理，计算企业多项数值指标，并进行一定的分析；最后，使用随机森林算法，对企业信用进行评级，并使用准确率指标评估算法性能。

**启发思考题：**

(1) 在本案例的数据预处理部分，我们为每个企业定义并计算了若干指标。考虑是否能找到一种合适的方法，可以自动地从原始数据集提炼有效信息，而非采用认为定义的方式。

(2) 除了案例中采用的决策树与随机森林算法，考虑引入其它机器学习算法或深度学习算法进行企业信用的评级，或对本案例采用的算法进行调优，从而取得更好的准确率。

**建议课堂计划：**

1. 时间安排：3课时
2. 学生学习准备：了解常见的数据预处理、决策树分类算法、集成学习算法的理论知识。
3. 分组及讨论内容：分组讨论启发思考题。
4. 案例开场白：在金融行业中，对银行来说选择给哪家企业发放贷款是一个重要的问题，这一选择通常由银行负责相关方面的经理根据经验进行主观判断。但是人工评判可能存在效率低、易受人为因素干扰的缺点，因此采用计算机与算法的方式进行企业信用评级是很有价值的一项研究。在本案例中。我们将使用大数据与机器学习的相关方法对企业的信用等级进行评估，以帮助银行进行合理地选择企业发放贷款。
5. 结束总结：该案例的选取金融贷款中企业多指标信用评估为主要问题，使用123家企业的多维度数据，结合数据挖掘、机器学习的相关方法，进行企业信用等级的评估。需要注意的是，课程所学内容是有限的，希望对该领域感兴趣的同学可以自行搜集相关资料，进一步了解大数据在金融风控领域的其它应用及相关技术。
6. 案例引导建议：在教师简要分析应用场景后，鼓励学生自主考虑如何使用所学知识解决该场景问题。

**参考文献：**

[1] Del Río S, López V, Benítez J M, et al. On the use of mapreduce for imbalanced big data using random forest[J]. Information Sciences, 2014, 285: 112-137.

[2] 王强.决策树过拟合问题研究[D].安徽：合肥工业大学,2008.

[3] 周志华. 机器学习[M]. 北京: 清华大学出版社, 2016.

**其他教学支持材料：**

1. 本案例提供配套的PPT、数据集与代码等，发布于Github，链接为：https://github.com/Wanghui-Huang/CQU\_bigdata。