# 股票大数据分析平台的构建案例教学指导手册

**教学适用的课程：**《大数据架构与设计》

**教学适用的专业：**电子信息

**教学目的与用途：**理解股票大数据分析平台的构建；掌握大数据预处理、数据分析相关知识点；掌握常见机器学习预测方法。

**教学内容：**

（1）理论依据：本案例基于中国股票大数据分析平台的构建。案例整合了常用的数据挖掘与机器学习方法，进行股票数据预处理及可视化，同时使用机器学习算法进行股票走势预测。

（2）涉及知识点：Hadoop；Spark；数据预处理；线性回归。

（3）分析路径：首先明确金融股票数据分析的意义，并以构建中国股票大数据分析平台为具体问题。通过Hadoop 和Spark搭建服务平台，此后操作均基于该平台。随后利用源于财经网站的股票数据构建数据集，并进行预处理及可视化分析。最后采用常用机器学习算法如线性回归进行股票走势预测。

**启发思考题：**

针对于股票走势预测，除了使用线性回归算法，还能想到什么其他方法吗？（参考答案：深度学习算法如LSTM、GRU等）

**建议课堂计划：**

1. 时间安排：3课时
2. 学生学习准备：了解Hadoop 、Spark以及常用机器学习算法。
3. 分组及讨论内容：分组讨论启发思考题。

案例开场白：随着大数据时代的到来，人们逐渐认识到了数据的重要性。数据不仅是一种资源，更是一种财富。在大数据应用领域中，金融数据分析被视为一个很有前景的方向。股票分析一直是金融领域一个很热门的话题，而且涉及多个领域的知识。在此之前，人们更多的是采用基本分析，即通过宏观及微观的经济政策、本行业领域的发展状况、投资者的行为态度、反映企业自身发展状况的指标等方面来预测股票今后走势。随着大数据相关技术的发展，在海量的股票历史数据中发现规律进而预测股票走势成为一个很热门的研究课题。

1. 结束总结：这节课我们以构建中国股票大数据分析平台为具体问题。通过Hadoop 和Spark搭建服务平台，此后操作均基于该平台。随后利用源于财经网站的股票数据构建数据集，并进行预处理及可视化分析。最后采用常用机器学习算法如线性回归进行股票走势预测构建电影信息推荐平台为例。我们通过平时课程所学知识实现了股票大数据分析平台的构建，课后感兴趣的同学可以搜集相关资料，进一步了解大数据在其他场景的应用，相信你们会收获更多。
2. 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] 周志华. 机器学习[M]. 北京: 清华大学出版社, 2016.

[2] 陈文颖. 基于Hadoop的股票数据聚类分析[D].浙江大学,2018.

**其他教学支持材料：**

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

2. 本案例涉及到数据预处理以及多种机器学习算法，建议使用python语言进行编写，推荐的工具包有Tushare（财经数据接口包）、scikit-learn（机器学习库）。