**体育大数据背景下的NBA球员分析教学指导手册**

**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：**理解体育大数据背景下的实际应用；掌握数据采集、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：本案例为基于体育大数据背景下的NBA球员综合素质评价。案例整合了常用的数据挖掘与机器学习方法，进行数据采集与数据、聚类及可视化，实现了大数据依托的运动员素质评估，为比赛指明了清晰的发展方向与策略。

（2）涉及知识点：体育大数据，数据采集，数据预处理，聚类分析，可视化。

（3）分析路径：首先介绍了数据支撑下的体育产业发展背景，爬取网站数据进行数据采集。通过数据预处理进行了特征删除、缺失数据处理以及数据上传操作。然后选取提供得分(PTS)、篮板(TRB)、助攻失误比(ATR)三个维度对球员进行聚类分析，通过绘制簇数量与误差平方和的图以寻找最优簇值K。最后将聚类结果进行高维映射并可视化展示。

**启发思考题:**

本案例中使用了Kmeans聚类算法，Kmeans方法有哪些缺点？有哪些改进方法？（参考答案：Kmeans聚类需要提前确定聚类的簇数量，而且受均值影响较大。改进方法比如均值漂移聚类法）

**建议课堂计划：**

1. 时间安排：3课时
2. 学生学习准备：了解常见的数据采集、数据预处理、聚类方法的理论知识。
3. 分组及讨论内容：分组讨论启发思考题。
4. 案例开场白：伴随着大数据时代的到来，数据被视为了一种新的生产要素和创新驱动力。“大数据”概念自提出以来，已迅速成为传统行业创新发展的重要参考。正值全民体育时代的今天，将体育与大数据融合已是大势所趋。这节课我们将讨论大数据背景下的球员分析场景。
5. 结束总结：这节课我们以大数据背景下的NBA球员分析为例，通过网络爬虫获取球员数据信息，并对数据进行缺失值填充和特征清洗；对球员数据进行聚类分析，从而评判球员综合素质；并对聚类分析结果进行可视化。相信大家通过本次课程案例的学习，感受到了体育产业决策背后的数据支撑，课下也欢迎同学们进一步深入了解大数据背景下的体育产业变革。
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] 周志华. 机器学习[M]. 北京: 清华大学出版社, 2016.

参考内容：数据预处理理论、Kmeans聚类算法

1. 吕云翔,张扬编.Python网络爬虫与数据采集[M].人民邮电出版社,2021.

参考内容：数据采集理论

**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pandas（数据读取与预处理库），matplotlib（可视化绘图库），PySpark的MLlib(机器学习算法库)。