# 地震数据分析与预测研究案例教学指导手册

**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：**理解实际问题的大数据应用；掌握大数据预处理和分析相关知识点；掌握大数据处理与分析的基本流程；掌握聚类算法的相关知识点；掌握神经网络的相关知识点。

**教学内容：**

（1）理论依据：本案例基于大数据技术分析和处理过往的地震数据。案例使用了常用的数据挖掘框架与机器学习方法，进行数据预处理，聚类分析。

（2）涉及知识点：数据挖掘；数据预处理；聚类分析；神经网络。

（3）分析路径：首先分析了地震的危害，介绍了对地震问题的分析在社会中的重要性。随后进行数据挖掘，得到过往的地震数据，并对挖掘到的数据进行预处理，主要包括数据清洗和可视化操作，以获得对数据的初步认识。然后使用Kmeans和DBSCAN聚类算法对地震数据进行聚类分析。最后使用神经网络预测地震震级，得出相应的结论。

**启发思考题：**

如果当前使用的算法对数据的处理效果差，如何分析并定位问题？（参考答案：可能是输入数据过少，算法没有学习到相应的知识；可能是选取的数据属性相关性低，可以考虑进行因子分析，找到与目标相关性更高的属性；可以适当发散思维）

**建议课堂计划：**

1. 时间安排：3课时
2. 学生学习准备：了解常见的数据挖掘相关知识、数据预处理相关方法、聚类分析的相关知识、神经网络的相关知识。
3. 分组及讨论内容：分组讨论启发思考题。
4. 案例开场白：几千年来，人类一直在与地震对抗。每次地震的发生，都将给人类社会带来惨重的灾难。如果我们能够预测地震的到来，提前做好应对准备，那么地震可能不会导致生命丧失和更大范围的财产损失。本节课，我们将从大数据的角度分析地震相关资料，从过去的地震数据中挖掘有用的信息，深入我们对地震的了解，让我们更好地认识地震这一自然灾害。
5. 结束总结：这节课我们分别爬取了地震的历史数据，对其进行预处理并可视化了处理结果；随后使用Kmeans和DBSCAN聚类算法来，聚类分析了地震发生的位置信息；最后，使用了神经网络模型来对地震震级进行预测。通过上述操作，同学们或许对地震这一现象有了更深入地认识。其实，在生活和科研中，人们都是通过总结过去的经验，来指导自己更好地应对未来遇到的各种问题。我们把这种行为称之为学习，这也是“机器学习”这一名称的由来。希望同学们在掌握了这些知识后，也可以通过不断学习，来使自己也能应对以后发生的所有难题。
6. 案例引导建议：在教学过程中，教师可以通过分析案例中未考虑周到的因素，引导学生发散思维，寻找更多更好的方法来处理问题。

**参考文献：**

[1] 周志华. 机器学习[M]. 北京: 清华大学出版社, 2016.

[2] Goodfellow I, Bengio Y, Courville A. Deep learning[M]. MIT press, 2016.

**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scrapy（数据爬取），pandas（数据读取与预处理库），scikit-learn（机器学习算法库），keras（深度学习算法库）。