# 网易云音乐大数据系统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：**掌握大数据预处理、数据分析相关知识点；掌握数据可视化分析；掌握常用机器学习、深度学习方法；掌握图数据库的使用。

**教学内容：**

（1）理论依据：本案例基于网易云音乐大数据系统的构建。案例整合了常用的数据挖掘与机器学习方法，进行网易云音乐官方网站的数据的爬取与存储，同时结合经典机器学习和深度学习方法进行评论情感分析以及对用户进行歌曲、歌手个性化推荐。

（2）涉及知识点：网络数据爬取；数据预处理与存储；机器学习；深度学习。

（3）分析路径：首先明确以构建网易云音乐大数据系统为具体问题。通过Hadoop 和Spark搭建服务平台，随后基于此平台采用网络爬虫进行网易云音乐官方网站的数据爬取和存储，利用构建的数据集结合SVM、贝叶斯模型、LSTM以及BERT等模型进行评论情感分析，最终基于图结构数据对用户进行歌曲、歌手推荐。

**启发思考题：**

除了利用SVM、LSTM和BERT等模型进行评论情感分析，还可以使用什么模型方法？（参考答案：请查阅相关文献，检索目前先进的方法模型）

**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、网络爬虫以及常用机器学习、深度学习方法理论知识。
3. 分组及讨论内容：分组讨论启发思考题。

案例开场白：近年来各种音乐软件层出不穷，网易云音乐作为国内主流的在线音乐提供平台之一，拥有大量的入驻歌手、音乐资源与活跃用户，既为用户提供丰富多样的音乐资源，又发展出了极具特色的评论区，产生了大量的社交数据，深受青年用户的喜爱，并获得了“云村故事会”、“网抑云”等特色标签。基于网易云音乐的海量音乐数据与社交数据，使用大数据分析与挖掘技术，可以得到以往的音乐流行趋势、歌手流行趋势、用户喜好等多个维度的信息，并在此基础上预测未来的发展情况。另外，基于网易云音乐的大量评论文本数据，使用自然语言处理技术，可以分析用户评论所包含的情感信息及“抑郁”属性，深入了解当代青年人的心理状态与需求，对广告个性化投放、情感类产品研发等实际应用场景都有重要价值，亦可以结合音乐大数据为用户进行歌曲推荐、歌手推荐、好友推荐等。因此，对网易云音乐中的大数据进行分析，是非常有研究价值的一个课题。

1. 结束总结：这节课我们以构建网易云音乐大数据系统为具体问题。通过Hadoop 和Spark搭建服务平台，随后基于此平台采用网络爬虫进行网易云音乐官方网站的数据爬取和存储，利用构建的数据集结合SVM、贝叶斯模型、LSTM以及BERT等模型进行评论情感分析，最终基于图结构数据对用户进行歌曲、歌手推荐。我们通过平时课程所学知识实现了基于网易云音乐的大数据系统，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] 阮泽楠. 音乐社交平台用户情绪特征研究[D]. 浙江理工大学,2019.

**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requests（网络请求库），BeautifulSoup（HTML解析库），scikit-learn（机器学习库）。