《金融大数据分析》教学大纲

课程编号：2121503B

课程类型：□通识教育必修课 □通识教育选修课

□学科基础课 □专业核心课

☑专业提升课 □专业拓展课

总学时：48 讲课学时：32 实验（上机）学时：16

学　　分：3

考试类型：□考试 ☑考查

适用对象：信息管理与信息系统专业

□是 ☑否 适合作为其他专业学生的个性化选修课

先修课程：量化分析方法与实践、投资学、数据分析理论与实践(双语)、金融学、统计学

一、教学目标

1、专业教育目标

在金融科技发展的市场背景下，对金融大数据的挖掘、分析和利用已经成为金融管理过程中的重要工具和手段。本课程主要学习金融大数据的相关原理、分析方法和技术，并根据实际业务需求选择大数据分析工具（Python）对不同类型的数据进行处理和分析，最终应用于金融管理实践当中。

通过本课程学习，学生能够了解金融大数据的主要来源和特征，熟练掌握金融大数据的采集、传输、处理、分析和应用的相关技术，掌握Python的基础语言知识、大数据分析和应用技术，能够基于Python进行数据采集、处理、分析和应用工作，能够对金融市场中的风险管理、投资者情绪等典型问题展开实证分析。鉴于金融大数据分析已经成为金融管理过程中的重要工具和手段，因此该课程是信息管理与信息系统培养方案的重要组成部分，是先修课程“统计学”、“数据分析理论与实践”和“量化分析方法与实践”的实战延续，并可为同期进行的专业必修课“投资者风险管理”打下良好技术基础。

2、课程思政目标

习近平总书记指出要实施国家大数据战略，加快建设数字中国，并多次强调了推动金融业高质量发展的重要性。本课程以习近平总书记的讲话精神为引领，以金融大数据分析的专业教育为基础，实现以下几方面的思政教育目标：

（1）使学生认识到数据是金融领域的新生产要素，领会金融及金融数据安全的重要性。穿插中美贸易摩擦对中国大数据的技术限制，使学生领会习总书记讲话的必要性和正确性，激发学生爱党爱国热情。

（2）使学生认识到中国特色社会主义背景下金融大数据存在的时代、行业和学科特征，为树立正确的价值观和职业道德观奠定基础。

（3）使学生认识到掌握金融大数据分析技能的紧迫性，学会用大数据技术分析金融行业发展过程中存在的金融风险等问题，为金融行业的风险监管和行业健康发展做出贡献。

二、教学内容及其与毕业要求的对应关系

（一）教学内容

本课程主要教学内容是在学生掌握基本的统计学、数据分析、Python实际操作等基本原理和算法后，根据金融大数据的特征，展开一系列背景分析、模型导入、Python应用工作，并应用于实际金融市场典型问题的分析当中。具体包括：金融大数据来源和特征分析，Python在金融市场中的应用，金融大数据探索、数据预处理、挖掘建模、案例分析等。

（二）教学方法和手段

根据教学目标，拟采用的教学方法有：课堂讲解基本概念和核心知识，讲授和讨论相结合领会知识要点，案例教学训练解决问题的能力，最后借助Python开发环境软件让学生进行上机操作和具体实践。

（三）实践教学环节要求

根据教学进度和要求布置相应的小作业，通过上机实践加深对理论知识的理解与吸收。建议在先行课的基础上，提前安装和配置Python软件，尝试完成书上的案例内容。通过上机学习本课程的理论，掌握基于Python语言和工具的金融大数据分析方法。

（四）学习要求

为有效学习本课程，要求学生首先具备数据库、面向对象程序设计与开发等方面的基本知识，熟练掌握面向数理统计与数据挖掘等方面的完整专业知识。

自学时应该熟读大纲，提纲挈领地掌握金融大数据的分类和特征、Python数据分析的内容，随后按照大纲熟读教材，并通过课后思考和上机实践进行多角度和多层次的反复学习。

（五）与毕业要求的关系

随着大数据时代的到来，大型企业、政府和机构逐渐积累了大量的业务数据，基于互联网技术和大数据分析手段挖掘金融大数据的价值，逐渐成为金融行业竞争的必要组成部分，也是未来发展的必然趋势。因此，学习和掌握金融大数据的分析方法、技术及相关应用，不仅是当前金融行业发展的需要，还是适应金融行业及其它企业运营和竞争的基本要求，也是信息管理与信息系统专业培养的基本要求。

（六）教学中应注意的问题

由于数理统计原理和数据挖掘技术原理是该课程的先修基础，如果学生先修基础课没有学好，在运用Python语言进行数据分析时，则会出现不知道如何选用并且不理解数据分析结果的问题。因此，教学中需要根据学生掌握先修课程基础情况，由易到难循序渐进学习数据分析相关工具和技术。

三、各教学环节学时分配

**教学课时分配**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 序号 | 章节内容 | 讲课 | 实验 | 其他 | 合计 |
| 第一章 | 金融大数据概述 | 8 |  |  | 8 |
| 第二章 | Python在金融大数据分析中的应用现状及前景分析 | 6 | 2 |  | 8 |
| 第三章 | 数据获取和预处理 | 4 | 4 |  | 8 |
| 第四章 | 数据挖掘建模 | 6 | 4 |  | 10 |
| 第五章 | 案例1：金融投资者情绪分析 | 4 | 4 |  | 8 |
| 第六章 | 案例2：金融行业风险识别 | 4 | 2 |  | 6 |
| **合计** | | **32** | **16** |  | **48** |

四、教学内容

第一章 金融大数据概述

课程思政切入点：结合金融科技、金融安全、数据分析特征，讲述习近平总书记关于金融大数据的讲话和指示。穿插中美贸易摩擦对中国大数据技术的限制等事件，使学生领会习总书记讲话的必要性和正确性，激发学生的爱党爱国热情，以及对本课程的学习热情。

教学内容：概要介绍在金融科技发展的背景下，商业银行、保险公司等传统金融行业与第三方支付、网络众筹等新兴金融业态的发展情况，探索金融大数据的来源和特征并分析研究金融大数据的必要性。

重点和难点：金融行业特征，大数据来源、分类和获取方式，应用领域。

考核要求：使学生了解新时代背景下金融行业所包含的业态及相互关系，总结金融大数据的来源、特点和特征，深入理解运用金融大数据分析金融问题的必要性和重要性。

第二章 Python在金融大数据分析中的应用现状及前景分析

课程思政切入点：介绍金融大数据存在的合理性和必要性及其对金融行业发展的影响。将金融大数据分析场景和Python工具结合，引用习近平总书记在纪念马克思诞辰200周年大会上发表重要讲话时指出的“学习马克思,就要学习和实践马克思主义关于生产力和生产关系的思想”，使用马克思主义辩证法，引导学生辩证地认识Python工具。

教学内容：Python在金融大数据分析中的现状分析、Python复习、必要安装包的使用。

重点和难点：Python在金融大数据应用总结分类、Python操作复习。

考核要求：了解Python在金融大数据分析中的应用现状和前景，Python常用数据分析工具的种类、安装配置与功能定位，掌握Numpy、Matplotlib和Pandas类库的编程接口与使用方式，能够应用上述类库进行初步的数据统计。

实验一：给定样本数据，练习使用Numpy和Pandas进行基本的统计分析。

实验二：给定样本数据，练习使用Matplotlib进行基本的统计绘图。

第三章 数据获取和预处理

课程思政切入点：数据获取和预处理是金融大数据分析的基础。运用辩证唯物主义认识论引导学生认识数据获取和数据预处理的意义。认识具有反复性，由于受主客观条件的限制，数据获取和预处理的过程不是一帆风顺的，这就决定了学生对数据获取和预处理过程的正确认识往往要经过从实践到认识、再从认识到实践的多次反复。只有这样，学生才能更好更快地掌握数据处理技能。

教学内容：数据爬取，数据质量分析，数据特征分析，数据预处理中的清洗、集成、交换等。

重点和难点：Python主要数据探索函数的使用与结果分析、预处理函数的使用。

考核要求：了解数据质量分析和数据特征分析常用方法，熟练掌握Python主要数据探索函数的使用方法，理解数据探索分析结果；了解数据清洗、数据集成、数据变换的基本概念和原理，掌握Python主要数据预处理函数的使用。

实验三：股票评论数据的数据探索分析。

实验四：股票评论数据的数据预处理。

第四章 数据挖掘建模

教学内容：主题识别，聚类分析，关联规则，情感分析。

重点和难点：基于Python的文本主题识别、情感分析相关算法和工具使用。

考核要求：了解Python的聚类分析、主题识别、情感分析的相关算法，掌握基相关包的使用，理解数据分析结果。

实验五：股票评论数据的主题识别和情感分析。

实验六：从财务报表中识别企业风险因子。

第五章 案例分析1：金融投资者情绪分析

课程思政切入点：金融体系与实体经济中的企业息息相关，证券价格的波动会影响实体经济发展，进而影响到社会的长治久安。而投资者作为证券市场的重要组成部分，投资者情绪的波动将会影响到投资者的决策，从而影响证券市场的发展。因此，要深入分析投资者情绪的影响因素及其对证券价格的影响机制，为我国证券市场的健康发展保驾护航。

教学内容：综合利用已学数据分析和数据挖掘知识，运用Python从股票评论数据中识别不同板块投资者的主题演化，构建投资者情绪，并分析投资者情绪与股价的关系。

重点和难点：数据分析各阶段工具的综合利用与结果解释。

考核要求：能够综合运用Python提供的各类分析工具，对给定的金融数据进行必要的分析，并能够理解分析结果。

实验七：股票评论数据的综合分析。

第六章 案例分析2：金融行业风险识别

课程思政切入点：金融风险是影响金融行业发展的关键因素，而风险识别是金融风险管理的基础。在教学过程中，从近年来银行、证券、保险、及P2P和第三方支付等金融子领域暴露的风险事件出发，分析金融风险识别的重要性以及对金融行业风险监管和发展的影响，进而为金融监管提供理论和数据支撑。

教学内容：综合利用已学数据分析和数据挖掘知识，运用Python从财务报表中识别企业风险因子，并分析风险因子与企业股价的关系。

重点和难点：数据分析各阶段工具的综合利用与结果解释。

考核要求：能够综合运用Python提供的各类分析工具，对给定的金融数据进行必要的分析，并能够理解分析结果。

实验八：股票评论数据的综合分析。

五、考核方式、成绩评定

本课程采用平时、期末作业相结合的考核方式，尤其注重实践环节的考核。各部分所占比重为平时成绩占40%，期末大作业成绩占60%。

平时成绩包括考勤成绩（10%）和上机作业成绩（30）。

期末考核采用大作业形式，指定某个具体问题，要求学生能够对数据进行处理、编码和分析，最后提供相应的数据分析方案和分析报告。考核方式注重知识应用能力和解决问题能力的考核，需要学生提交自己开发的基于Python的数据分析方案及分析报告。

六、主要参考书及其他内容

（一）推荐教材

1. Hilpisch Y著，姚军译. Python金融大数据分析. 第二版. 北京：人民邮件出版社，2020，ISBN：9787115521330.

2. 王宇韬，房宇亮，肖金鑫, 等著. Python金融大数据挖掘与分析全流程详解. 北京：机械工业出版社，2019，ISBN：9787111633204.

（二）主要学习参考书

1．Nelli F 著；杜春晓 译. Python数据分析实战. 第二版. 北京：人民邮电出版社, 2019. ISBN：9787115522023.

2． 张良均，谭立云，刘名军 等著. Python数据分析与挖掘实战. 第二版.北京：机械工业出版社，2019. ISBN：9787111640028.

3. 王青天著. Python金融大数据风控建模实战：基于机器学习. 北京：机械工业出版社，2020，ISBN：9787111655794.

执笔人：姚银红 教研室主任：胡磊 系教学主任审核签名：