



数据科学在大数据金融  
行业中的应用

Data Science and its Application in Finance and Big Data



#1. 课程背景及简介



数据科学正在以空前的速度发展,我们日常的生活也越来越趋近于数字化。据估计,在过去的两年中,地球上的每一个人每秒会产生 170 兆的数据。因此,数据科学已成为 21 世纪最热门的学科之一。其中数据科学在商业领域的应用发展最为迅猛,世界各国的公司和组织都在不遗余力地从海量数据中提取和分析有意义的内容,以做出更明智的商业决策。本课程正是这样一门数据科学与智能金融交叉的课题,学生将通过数据分析,帮助公司做出正确的金融决策。

本课程将系统讲解数据科学的理论和应用,使学生能够广泛的了解全球背景下第四次工业革命 市场的技术增长动态。旨在培养多元化的商业人才,推动以研究为基础的创新,从而促进技术的可持续增长。此外,本课程还将帮助学生熟悉影响全球经济发展的经济、社会、政治和技术等原因,提高参与全球商业环境所需的数据分析能力。

这门数据科学与智能金融课程将讨论数据科学中一些最热门的主题,包括理论和实践两方面。理论方面包括数据库、数据仓库、数据分析、数据科学、数据挖掘等方面的技能。实践方面学生将有机会利用模拟数据和数据分析软件解决真实世界的商业问题。同时还将了解时间序列分析,神经网络,推荐引擎,回归,计算机视觉,人工智能等前沿技术。

#2. 学习目标



本课程将解决许多挑战, 如:

- ★ 学习数据科学的基本概念。
- ★ 掌握共享信息, 数据采集, 数据的遵从性和安全性
- ★ 介绍使用数据分析软件的实用技能
- ★ 探索跨不同行业的真实案例并掌握合适的数据挖掘方法
- ★ 分析商业智能的未来趋势及其对组织和社会的影响
- ★ 熟悉复杂的数据科学技术及其应用以解决现实问题, 促进数据驱动的决策
- ★ 提升数据分析与提取的能力
- ★ 了解设计交互式数据可视化评估的道德和隐私问题

#3. 任课教师信息



Dr. M M

目前担任加州大学河滨分校商学院和波莫纳加州州立理工大学讲师,曾任美国亚洲基金会项目经理和克莱蒙特研究生大学信息技术主任。是数据科学领域的专家, 超过 20 余年的相关领域工作与执教经验。除此以外, Moniruzzaman 博士还在公共、私人、非政府双边或多边国际发展组织有丰富经验。

# 4. 课程设置

PBL

周期	时间	课程设置内容	课时
第一周 学习指南 教授及助教辅导	1月28日 周六	什么是PBL教学方法	1
		PBL教学的常见形式	1
	1月29日 周日	教授课-1 交叉学科PBL课程设计及知识点学习 学习目标：数据科学 描述：数据科学的基础理论，数据预处理，数据计算与数据管理。数据科学对计算机科学的影响。数据科学所涉及的技术，包括聚类降维与机器学习。	3
	1月30日 周一	助教课-1 知识点查漏补缺	2
	1月31日 周二	教授课-2 制定项目方向 学习目标：商业分析与可视化表达 描述：数据可视化的展现形式，数据可视化的作用，复杂交互式图表可视化；Tableau构建基本图表、高阶图表和数据仪表盘；共享信息、数据采集、数据遵从性和安全性、数据共享、商业智能、数字仪表板的战略价值。	3
	2月1日 周三	助教课-2 知识点查漏补缺	2
	2月2日 周四	教授课-3 交叉学科课程知识点学习 学习目标：数据仓库和数据挖掘/数据科学 描述：数据仓库的商业应用，数据仓库的体系结构；OLAP技术；数据挖掘过程，数据挖掘方法（决策树，神经网络，概率论与数理统计），聚类分析	3
	2月3日	助教课-3	2

第二周 教授及助教辅导	周五	知识点查漏补缺&跟进项目调研进度	
	2月4日 周六	教授课-4 互动与项目设计跟进答疑	1.5
	2月6日 周一	助教课-4 跟进项目调研进度	2
	2月7日 周二	教授课-5 交叉学科课程知识点学习 学习目标：文本挖掘；网络分析； 社交媒体分析；大数据；物联网； 云计算 描述：文本预处理，文本分类， 文本聚类，网络内容，结构，使用挖掘。 网络分析在搜索引擎，电子商务等领域中的应用；云计算的服务类型与体系结构	2
第三周 教授及助教辅导 未来展望	2月8日 周三	助教课-5 跟进项目调研进度	2
	2月9日 周四	教授课-6 交叉学科课程知识点学习 学习目标：机器学习；未来的方向；人工智能 描述：人工智能的发展方向与机遇，机器学习未来热点研究	2
	2月10日 周五	助教课-6 知识点查漏补缺&指导项目成果展示	2
	2月11日 周六	教授课-7 教授点评项目成果	1.5
	2月12日 周日	升学与就业方向展望	1
		个人规划及发展建议	1
总课时	32		

#5. 阅读材料

PBL

Journals:

★ Data Scientist: The Sexiest Job of the 21st Century  
<https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century>

★ AutoODE: Bridging physics-based and data-driven modeling for COVID-19 forecasting  
<https://www.amazon.science/publications/autoode-bridging-physics-based-a>



nd-data-driven-modeling-for-covid-19-forecasting

★ How AI is changing the video game industry: augmentation and synthetic media  
[https://aibusiness.com/author.asp?section\\_id=789&doc\\_id=761220](https://aibusiness.com/author.asp?section_id=789&doc_id=761220)

★ A Look into Microsoft's Data-Driven Approach to Improving Sales  
<https://hbr.org/2018/12/a-look-into-microsofts-data-driven-approach-to-improving-sales>

**Textbooks:**

★ Camm, J., et al. (2019). Essentials of business analytics (3rd ed). Boston, MA: Cengage Learning.

★ Nussbaumer Knafllic, C. (2015). Storytelling with data: A data visualization guide for business professionals. Hoboken, NJ: John Wiley & Sons.

★ Shearer, C. (2000). The CRISP-DM Model: The new blueprint for data mining Journal of Data Warehousing, (5)4. pp. 13-22.

★ Berry, M. and Linoff, G. (2011). Data mining techniques for marketing, sales and customer relationship management, (3rd ed.). John Wiley & Sons, New York, NY.

#6.项目主题

PBL

本课程使用 PBL 教学法，PBL 即项目式学习，是一种以学生为中心的教学方法，教师提供关键素材构建学习环境，学生通过在此环境里解决一个开放式项目的经历来学习。以下为本课程可选的项目主题：

- 如何根据商业需求筛选和收集数据
- 如何规划数据建模来解决商业问题
- 如何利用数据分析软件做出最优商业决策
- 如何设计数据交互式可视化仪表盘
- 如何评估道德和隐私问题对使用商业智能的影响

英文版教学大纲

PBL

Course Title	Data Science and its Application in Finance and Big Data
Credit Hours	32 (one credit hour is 45 minutes)
Course Objectives	This class will apply a project-based approach to learning and encourages the pragmatic application of a variety of tools and methods to solve complex business problems. After completing this class, you will acquire the following skills: Data Science Life Cycle



	<p>Business Intelligence</p> <p>Data Analytics &amp; Visualization</p> <p>Data Science</p> <p>Data Warehousing</p> <p>Data Mining (Statistical Analysis and Exploring)</p> <p>Machine Learning</p> <p>Future Trends</p> <p>Ethics and Privacy</p> <p>You will complete a capstone project by executing a culminating project that integrates the core skills and concepts learned throughout the course. The capstone combines the technical, analytical, interpretive, and social dimensions required to design and execute a full data science project. You will learn integral skills that prepare you for long-term professional success in the data science field.</p>
Course Description	<p>This business-oriented data science course will discuss some of the most important topics in data science, covering both theoretical aspects and hands-on projects. This class will facilitate the opportunity for gaining an eclectic understanding of the state-of-the-art data science concepts, models, and techniques. This class will familiarize you with a broad cross-section of models, algorithms and software for data warehousing, data mining, machine learning, text-mining, AI, and prepare you for real world problem simulation in data science projects.</p> <p>In this program that lasts for 10-hour online sessions (5 Modules) with 2-hour Q&amp;A session, you will be able to upgrade your skills in data spectrum consists of database, data warehouse, data analytics, data science, data mining &amp; machine learning concepts by learning the theory and practical application of supervised and unsupervised learning, time-series analysis, neural networks, recommendation engines, regression, computer vision, artificial intelligence and to name a few. The program will be also equipped with a combination of conceptual and applied business topics with frequent lab sessions. The program arranges live and personalized mentored learning sessions by Data Science and Machine Learning experts to coach you to practically work on the industry relevant projects and get hands-on understanding. Tools to be used in the class include Tableau,</p>

	Excel (Pivot Table), Rapid Miner, Google Analytics and Microsoft Azure Machine Learning Tool.
--	---

**The topic in the global context**

Data is being generated at an unprecedented pace, gathered from all corners of our lives. It is estimated that for every person on earth, 1.7 MB of data have been created every second each day in the last two years. Thus, Data science has emerged as one of the defining disciplines of the 21st century. To be an internationally recognized leader in business education and research, organizations are sparing no effort to extract meaningful insights to make smarter business decisions from the ever-growing expanse of data that is being collected today.

**Brief introduction of the course**

In this highly interactive 12-hour class, Professor Moniruzzaman will introduce basic and applied data science methods and approaches that prepare our students to be effectively aware and fit enough for the Fourth Industrial Revolution labor market with an eclectic understanding of the technological dynamics of growth in both regional and global context. In this regard, Prof. Moniruzzaman will combine individual assessments and feedback profiles with presentations, interactive exercises, and discussion to support participants in enhancing understanding of the evolution of data spectrum and several major topics in data science.

This course is primarily designed to develop diverse business leaders, propel research-based innovation, and promote the sustainable growth. In addition, this course will facilitate the students to be conversant with dominant economic, social, political, and technological trends and conditions influencing foreign investment and development of the global economy and demonstrate the features of the cultural, interpersonal, and analytical competencies required for engaging in global business milieu.

	Topics
Module 1	Objective: Data Science Description: Introduce basic data spectrum and data science terminology; Data pre-processing, data calculation and data management. The impact of data science on computer science. The technologies involved in data science include clustering dimension reduction and machine learning

Module 2	Objective: Business Analytics & Visualizations Description: To share strategic value of information, data acquisition, data compliance and security, data sharing, data visualization, business intelligence, data reporting, digital dashboards. To design interactive visualizations of data for personalized dashboards.
Module 3	Objective: Data Warehousing & Data Mining/Data science Description: Introduce hands on marketable skills in using data analytics software. Apply data mining algorithms to extract novel, valid, and understandable patterns from data.
Module 4	Objective: Text mining; Web analytics; Social media analytics; Big data; IoT; Cloud computing Description: Text pre-processing, text classification, and text clustering. Use of mining network analysis in search engines, e-commerce and other fields of application. Service types and architecture of cloud computing
Module 5	Objective: Hadoop; Machine Learning; Future Direction; AI Description: The development direction and opportunity of artificial intelligence. The hot research of machine learning in the future.

Required Readings

Journals

Data Scientist: The Sexiest Job of the 21st Century  
<https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century>

AutoODE: Bridging physics-based and data-driven modeling for COVID-19 forecasting  
<https://www.amazon.science/publications/autoode-bridging-physics-based-and-data-driven-modeling-for-covid-19-forecasting>

How AI is changing the video game industry: augmentation and synthetic media  
[https://aibusiness.com/author.asp?section\\_id=789&doc\\_id=761220](https://aibusiness.com/author.asp?section_id=789&doc_id=761220)

A Look into Microsoft's Data-Driven Approach to Improving Sales  
<https://hbr.org/2018/12/a-look-into-microsofts-data-driven-approach-to-improving-sales>

Textbooks

Camm, J. , et al. (2019). Essentials of business analytics (3rd ed). Boston, MA : Cengage Learning.

Nussbaumer Knaflic, C. (2015). Storytelling with data: A data visualization guide for business professionals. Hoboken, NJ: John Wiley & Sons.





Shearer, C. (2000). The CRISP–DM Model: The new blueprint for data mining Journal of Data Warehousing, (5)4. pp. 13–22.

Berry, M. and Linoff, G. (2011). Data mining techniques for marketing, sales and customer relationship management, (3rd ed.). John Wiley & Sons, New York, NY.

#### **Suggested list of the topics for the final project**

How to collect data sources and organize data to fit your content needs?

How to explore new insights from existing datasets?

How to plan a data science project?

How to plan data modeling to solve business problems?

How to use data analytics software to help decision making?

How to design interactive visualizations of data for personalized dashboards?

How to assess the impact of ethical and privacy issues on the use of business intelligence in business settings?

#### **A word for my students**

“The goal is to turn data into information, and Information into insight” Carly Fiorina, Ex–CEO of Hewlett Packard.

Are you one of the many who dreams of becoming a data scientist? In this course I will show you how does it really work under the hood! We will learn data science based on real–life projects and understand how exactly it solves real industry challenges. Welcome to this interesting course on data science and I'm looking forwarding to meeting you soon!