**Institute of Data Science**

**Graduate Program**

**RE50200 Data Science (資料科學)**

**Fall 2023 (112學年度第1學期)**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1. ***The mission of the College*** *is to serve business and society in the global economy through developing professionally qualified and socially responsible business leaders as well as through advancing the frontiers of knowledge in business management.* 2. ***The strategic objective of*** *Institute of Data Science*is to cultivate quality professionals with enthusiasm and global perspectives.   **Graduate Program Learning Goals** (goals covered by this course are indicated by checks):   |  |  |  | | --- | --- | --- | | ✓ | 1 | Graduate students should be able to appreciate data analysis approaches and to present research findings/ results effectively in speaking and in writing. | | ✓ | 2 | Graduate students should be able to integrate different functional areas in solving data analysis problems. | | Graduate students should be able to analyze data effectively and to recommend effective statistical methods. | |  | 3 | Graduate students should be able to demonstrate leadership skills as a data analysis team leader. | | Graduate students should be able to identify ethical dilemmas and to determine necessary courses of action. | |  | 4 | Graduate students should possess a global data science perspective and an awareness of the global business. | | ✓ | 5 | Graduate students should be able to coordinate actions and solve problems jointly with other members of a professional team. | |

**Instructor/開課教師:**

|  |
| --- |
| Shuen Lin Jeng/ 鄭順林  sljeng@mail.ncku.edu.tw  (06)2757575#53640 |

**Prerequisite/先修科目:**

程式設計Programming language、計算機概論 Introduction to computers

**Course Description/課程描述:**

本課程旨在介紹資料科學的導論，系統性地教授資料科學的核心基本技術與原理，包含資料收集、資料儲存、資料呈現、資料分析方法。本課程規劃介紹主要的統計分析方法和機器學習的演算法。除了講解資料科學的基本概念外，並配合不同的應用方向和程式撰寫，讓學生能實際體驗其相關所需之知識及技術能力。

The aim of the course is to provide an introduction to data science. The course will teach the core techniques and theory of data science by explaining the basic concept of data science and by using the techniques for various applications with programming skills. The content includes data collection, data storage, data display, and methods of data analysis. The course will provide the major statistical analysis methods and machine learning algorithms. The students will obtain the required methods and knowledge through practical training.

**Course Objectives/課程目標:**

1. 使同學對資料科學有基本了解。學習資料收集、資料儲存、資料呈現、資料分析方法。

Students will learn the basic knowledge of data science including the principal methods of data collection, data storage, data display, and data analysis.

1. 同學們能學會統計分析和機器學習的主要方法。

Students will comprehend the major methods in statistical analysis and machine learning.

1. 培養同學具備使用程式進行資料分析的基本素養。

Students will learn the programming skills for data analysis

**Content Summary/課程內容:**

|  |  |
| --- | --- |
| **Week** | **Sub-topic** |
| 1 | Introduction to Data Science: Data Source, Type, and Storage |
| 2 | Methods for Data Collection I: Design of Experiments and Survey Sampling |
| 3 | Data Manipulation |
| 4 | Data Visualization  Project Discussion |
| 5 | Working with Large Datasets  Project Discussion |
| 6 | Statistical Learning  Project Proposal |
| 7 | Linear Regression |
| 8 | Classification |
| 9 | Resampling Methods |
| 10 | Linear Model Selection and Regularization |
| 11 | Moving Beyond Linearity  Project Presentation |
| 12 | Machine Learning |
| 13 | Survival Analysis |
| 14 | Multiple Testing |
| 15 | Methods for Data Collection II: Factorial and Fractional Factorial Design  Project Presentation |
| 16 | Methods for Data Collection III: Nested design and Response Surface Methodology, |
| 17 | Final Project Presentation |
| 18 | Final Project Presentation |

** Teaching Approach(es)/教學方法 (如講授、討論、實作、報告、參訪、影音欣賞、個案研討、其他等)**

|  |  |  |
| --- | --- | --- |
| **序號** | **教學方法teaching method** | **分配(%)** |
| **1** | **講授 lecture** | **60%** |
| **2** | **報告 student report** | **20%** |
| **3** | **作業實作 practice** | **20%** |
| **總計** | | **100%** |

**Textbook/教科書:**

1. James et al. (2021). An Introduction to Statistical Learning with Applications in R. Springer.
2. Zhang (2021) A Tour of Data Science Learn R and Python in Parallel.
3. Mailund (2017). Beginning Data Science in R.
4. Latpate et al. (2021). Advanced Sampling Methods. Springer.
5. Montgomery (2012). Design and Analysis of Experiments. 8th Edition. Wiley and Sons.

**Grading Policy/評量方式:**

1. **作業Assignments 40%**
2. **專題Projects 30%**
3. **期末考Final exam 30%**

**Grading Policy for AACSB Multiple Assessment:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 這是數據所的 | **作業Assignments**  40% | **專題**  **Project**  30% | **期末考Final exam** 30% |
| **COMMU** | ☑ Speaking |  | 25% |  |
| ☑ Writing | 40% | 25% | 40% |
| **CPSI** | ☑ Interdiscip. Competence/ Prob. Solving | 60% |  |  |
| ☑ Critical Thinking/ Innovation |  | 25% | 60% |
| **LEAD** | □ Leadership |  |  |  |
| □ Ethical Reasoning |  |  |  |
| **GLOB** | □ Global Vision |  |  |  |
| **VSP** | ☑ Teamwork |  | 25% |  |