|  |  |
| --- | --- |
| **School Responsible:** | School of Computer Science |

|  |
| --- |
| **Module Overview:** |
| Data mining refers to the process of deploying advanced analytical solutions throughout an organization, from initial planning to final implementation. This module will guide students through a typical life cycle, such as the CRISP-DM model, and examine each stage in detail, including the tasks and technologies involved.  This module covers a variety of data discovery techniques and algorithms can be used to identify patterns within the data. The main goal of the students in this module is to give an overview of the different steps in data mining such as business and data understanding, data preparation, modelling, evaluating, and deployment. We use both drag-and-drop modules software and some basic programming commands for creating basic pipelines. |

|  |  |
| --- | --- |
| **Learning Outcomes (LO):** | |
| On Completion of this module, the learner will be able to: | |
| **1** | Explain the role of data mining in an organisation through the different life-cycle stages. |
| **2** | Compare and contrast the main data mining techniques in supervised, unsupervised, and reinforcement learning. |
| **3** | Assess, combine, and construct data mining pipelines using visual software (e.g., Weka, Orange or similar) or basic coding instructions (e.g., R, Python, or similar). |
| **4** | Apply and evaluate various data mining techniques (prediction, clustering, feature selection…) for different problems. |
| **5** | Extract and communicate insights from the discovered patterns using data mining solutions. |
| **6** | Apply Machine Learning operations (MLOps) to generate efficient data mining applications |
| **7** | Assess and contrast various legal, ethical and management issues with projects |
|  |  |
|  |  |

|  |
| --- |
| **Indicative Syllabus:** |
| Module content will be broadly as follows:   * Overview   + Introduction to data mining and applications of data mining   + Data, Information, Knowledge   + Framing a business model   + How Data Mining fits within the organisation * Data Mining Life Cycle   + Stages of a DM project   + Explore various data mining life cycles.   + Evolving nature of roles and responsibilities of people involved in data mining projects. * Data Preparation   + Extracting and loading data mining   + Data transformations   + Data sampling   + Data aggregation   + Feature engineering * Exploring Data and Gaining Insights   + Using a variety of analytic methods to gain data insights   + Role of visualisations in pattern discovery   + Time-series forecasting   + Exploring and mining text * Data Mining Techniques. Explore the use of various techniques for structured and unstructured data including:   + Classification   + Regression   + Association rule analysis   + Data cluster analysis   + Anomaly Detection   + Reinforcement Learning * Understanding and evaluating the outputs and determine what to use. * Deploying Data Mining Solutions   + Issues around deployment of data mining solutions   + Combining multiple algorithms and models   + Creating pipelines for deployment   + Model management and when to retrain models and solutions. * Topics on the Management of the Data Mining Process and Life-Cycle   + Legal Issues   + Ethical Issues   + Biases in data   + Using and managing different technologies |

|  |  |
| --- | --- |
| **Learning and Teaching Methods:** | |
| Lectures, tutorials and computer laboratory sessions | |
| **Total Teaching Contact Hours** | 39 |
| **Total Self-Directed Learning Hours** | 61 |

|  |
| --- |
| **Module Delivery Duration:** |
| One Semester. |

|  |  |  |
| --- | --- | --- |
| **Assessment** | | |
| **Assessment Type** | **Weighting**  **(%)** | **LO Assessment (No.)** |
| In this exercise, participants will explore fundamental data mining techniques and apply them to a dataset, focusing on key aspects such as model selection, model evaluation, and result interpretation. | 20% | 2 to 5 |
| Undertake a Data Mining problem based on the topics covered in the first part of the module. To include problem definition, data exploration, identification of data insights, selection and use of appropriate algorithms, evaluation of results, and comparison of results with available research. | 30% | 2 to 6 |
| Written exam. | 50% | 1,2,4,5,6,7 |
|  |  |  |
| **Module Specific Assessment Arrangements (if applicable)** | | |
| 1. Derogations from General Assessment Regulations |  | |
| 1. Module Assessment Thresholds |  | |
| 1. Special Repeat Assessment Arrangements |  | |

|  |
| --- |
| **Essential Reading**: (author, date, title, publisher)   * Shmueli, G., Bruce, P. C., Yahav, I., Patel, N. R., & Lichtendahl Jr, K. C. (2017). *Data mining for business analytics: concepts, techniques, and applications in R*. John Wiley & Sons.   **Supplemental Reading**: (author, date, title, publisher)   * Han, J., Pei, J., & Tong, H. (2022). *Data mining: concepts and techniques*. Morgan kaufmann. * Linoff, G. S., & Berry, M. J. (2011). *Data mining techniques: for marketing, sales, and customer relationship management*. John Wiley & Sons. * Han, J., Kamber, M., & Pei, J. (2012). Data mining concepts and techniques third edition. *University of Illinois at Urbana-Champaign Micheline Kamber Jian Pei Simon Fraser University*. * Aggarwal, Charu C. *Data mining: the textbook*. Vol. 1. New York: springer, 2015.   Additionally, a range of supplemental reading will be used to support the topics being covered throughout the semester. These will include various websites, white pages, research papers, reoirts by industrial analysts, and reports, tutorials, etc from data and database management companies. |

|  |  |  |  |
| --- | --- | --- | --- |
| **Version No:** |  | **Amended By** |  |
| **Commencement Date** |  | **Associated Programme Codes** |  |

# Modules that are to be offered as Stand-Alone CPD Programmes must have an NFQ level assigned

\*Details of the assessment schedule should be contained in the student handbook for the programme stage.

**Date of Academic Council approval ………………………….**