# GE2324: THE ART AND SCIENCE OF DATA

**Effective Term**
Semester A 2022/23

## Part I Course Overview

**Course Title**
The Art and Science of Data

**Subject Code**
GE - Gateway Education
**Course Number**
2324

**Academic Unit**
Computer Science (CS)

**College/School**
College of Engineering (EG)

**Course Duration**
One Semester

**Credit Units**
3

**Level**
A1, A2 - Associate Degree
B1, B2, B3, B4 - Bachelor's Degree

**GE Area (Primary)**
Area 3 - Science and Technology

**Medium of Instruction**
English

**Medium of Assessment**
English

**Prerequisites**
Nil

**Precursors**
Nil

**Equivalent Courses**
Nil

**Exclusive Courses**
Nil

# Part II Course Details

**Abstract**

 "The future belongs to the companies and people that turn data into products". Data are related to every aspect of our life and we use data products every day. Some well-known examples include Google, Yahoo, Twitter, LinkedIn, Reddit, and Baidu. This course is to introduce interesting cases, findings, and techniques on the analysis of data.

The course will mainly focus on different aspects of data analysis from diverse disciplines such as geography, ecommerce, finance, medicine, and social networks.  Students will have chances to manipulate those data, and perform simple yet interesting discoveries.  Students will get familiarized with a number of computational tools; for instance, Infographics, Weka, Tanagra, R Language, Octave, and Gephi to manipulate, visualize, and understand data.  Very limited programming skills are required in the course.  Instead, students will be taught to think scientifically and creatively about data.  The aim is for students to not only analyse data, but also develop interesting applications out of the findings.

**Course Intended Learning Outcomes (CILOs)**

|   | CILOs | Weighting (if app.) | DEC-A1 | DEC-A2 | DEC-A3 |
|---|---|---|---|---|---|
| 1 | Describe the basic concepts and techniques for analyzing data from diverse disciplines. | | x | | |
| 2 | Use computational tools to manipulate, analyze, and visualize data, in order to infer interesting findings from the data. | | x | | |
| 3 | Create basic protocols for practical problems by combining the data analysis tools learnt and create data visualization for effective communication and presentations. | | | x | |
| 4 | Propose ideas for interesting applications for knowledge discovery from the data and design detailed steps to realize and verify those ideas. | | | x | |

A1: Attitude
Develop an attitude of discovery/innovation/creativity, as demonstrated by students possessing a strong sense of curiosity, asking questions actively, challenging assumptions or engaging in inquiry together with teachers.

A2: Ability
Develop the ability/skill needed to discover/innovate/create, as demonstrated by students possessing critical thinking skills to assess ideas, acquiring research skills, synthesizing knowledge across disciplines or applying academic knowledge to real-life problems.

A3: Accomplishments
Demonstrate accomplishment of discovery/innovation/creativity through producing /constructing creative works/new artefacts, effective solutions to real-life problems or new processes.

**Teaching and Learning Activities (TLAs)**

| | TLAs | Brief Description | CILO No. | Hours/week (if applicable) |
|---|---|---|---|---|
| 1 | Lectures | Lectures focus on the introduction of basic concepts of data and tools, knowledge discoveries and findings from the data, and the utilization of the tools to analyse and visualize data. | 1, 2, 3, 4 | 3 hours/week |
| 2 | Tutorials | Tutorial sessions allow students to get first-hand experiences on data and tools. They can also access tools of similar functions, to critically evaluate the suitability of each tool to the task at hand. | 1, 2, 3, 4 | 8 hours/semester |
| 3 | One Project | Students can apply the tools to study some new datasets. Also, students will choose suitable tools to create new solutions to verify simple yet interesting conjectures about our daily life activities. In addition, students are expected to work in group project to propose and develop data-based applications. For example, students who are working on social network data may study the network dynamics. Students who are working with geographical data may incorporate building information system. Students who are working with media data may distinguish between the direct and indirect correlations. | 2, 3, 4 | 2 hours/week |

| 4 | Two Assignments | Assignments allow students to review the technical content taught in the lectures and tutorials. They also have the chance to explore potential applications by applying the concepts taught in the lectures | 1, 2, 3 | 2 hours/week |
|---|---|---|---|---|

**Assessment Tasks / Activities (ATs)**

|  | ATs | CILO No. | Weighting (%) | Remarks (e.g. Parameter for GenAI use) |
|---|---|---|---|---|
| 1 | Assignments | 1, 2, 3 | 20 |  |
| 2 | Project | 1, 2, 3, 4 | 40 |  |

**Continuous Assessment (%)**

60

**Examination (%)**

40

**Examination Duration (Hours)**

2

**Additional Information for ATs**

For a student to pass the course, at least 30% of the maximum mark for the examination must be obtained.

**Assessment Rubrics (AR)**

**Assessment Task**

Assignments

**Criterion**

Answer questions precisely and concisely.

**Excellent (A+, A, A-)**

High

**Good (B+, B, B-)**

Significant

**Fair (C+, C, C-)**

Moderate

**Marginal (D)**

Basic

**Failure (F)**

Not even reaching marginal levels

---

**Assessment Task**

Project

**Criterion**

1. Collect data sets independently.
2. Apply the tools from the course and perform an interesting study.
3. Combine the tools to form new data analysis protocols.
4. Present and organize the results coherently.

**Excellent (A+, A, A-)**

High

**Good (B+, B, B-)**

Significant

**Fair (C+, C, C-)**

Moderate

**Marginal (D)**

Basic

**Failure (F)**

Not even reaching marginal levels

---

**Assessment Task**

Examination

**Criterion**

Answer questions precisely and concisely.

**Excellent (A+, A, A-)**

High

**Good (B+, B, B-)**

Significant

**Fair (C+, C, C-)**

Moderate

**Marginal (D)**

Basic

**Failure (F)**

Not even reaching marginal levels

---

# Part III Other Information

**Keyword Syllabus**

Data Science, Knowledge Discovery, Data Visualization, Data Mining, Data Correlation, Data Similarity, Data Clustering, Association Rule Discovery, and Network Analysis.

**Reading List**

**Compulsory Readings**

|   | Title |
|---|-------|
| 1 | Nil   |

**Additional Readings**

|   | Title |
|---|-------|
| 1 | Bramer, Max. Principles of Data Mining. Springer London, 2016. |
| 2 | Roiger, Richard J. Data Mining: a tutorial-based primer. CRC Press, 2017. |
| 3 | Han, Jiawei et al. Data Mining: concepts and techniques. Elsevier, 2011. |
| 4 | http://en.wikipedia.org/wiki/Data_science |
| 5 | http://radar.oreilly.com/2010/06/what-is-data-science.html |
| 6 | http://datascienc.es/ |
| 7 | http://www.cytoscape.org |

# Annex (for GE courses only)

**A. Please specify the Gateway Education Programme Intended Learning Outcomes (PILOs) that the course is aligned to and relate them to the CILOs stated in Part II, Section 2 of this form:**

Please indicate which CILO(s) is/are related to this PILO, if any (can be more than one CILOs in each PILO)

**PILO 1: Demonstrate the capacity for self-directed learning**

1, 2, 4

**PILO 2: Explain the basic methodologies and techniques of inquiry of the arts and humanities, social sciences, business, and science and technology**

1, 2, 3

**PILO 4: Interpret information and numerical data**

2, 3, 4

**PILO 5: Produce structured, well-organised and fluent text**

3, 4

**PILO 7: Demonstrate an ability to work effectively in a team**

3, 4

**B. Please select an assessment task for collecting evidence of student achievement for quality assurance purposes. Please retain at least one sample of student achievement across a period of three years.**

**Selected Assessment Task**

Several groups obtained interesting results from their projects.One group found the competitive patterns for tea and coffee trading. Another group analyzed the social network of a particular website. The third group studied the food networks using network analysis.