# University of Sunderland
# School of Computer Science

# CETM46 – Data Science Product Development
# 2022/23
# Sessional
# Assignment 2 of 2 – 60% of the summative value of the module

The module will be assessed by two coursework assignments, which will cover related learning outcomes of the module respectively. In this second assignment (2 of 2), students will develop a data science product prototype and write up a report about the design, development and management of the prototype development project.

**ASSIGNMENT TITLE**

***A Domain Specific Data Science Product Development Project: Product Prototype and Project Report***

**LEARNING OUTCOMES ASSESSED**

**Knowledge**

K2. Utilisation of the state of art of data science methodologies and software tools for data analysis applications development

K3. Critical understanding of modern data science systems and their ecosphere

**Skills**

S2. Ability to design and develop data science systems using various data repositories and data models

S3. Developing data science products with modern data systems, visualisation technologies, software tools and their ecosphere

**ASSIGNMENT INTRODUCTION**

As a data scientist at a data science start-up company, you are tasked to design and develop a bespoke data science product for a specific application domain as part of an individual R&D project. The end users of your data science product have no or very limited knowledge of data science technologies, but expect to install, deploy and use your data science product for their company or organisation in an easy and user-friendly style.

The design and development of the product and the management of the R&D project will be based the following assumptions:

1) The data source MUST be from opensource datasets, such as https://data.gov.uk/, the UCI Machine Learning Repository https://archive.ics.uci.edu/ml/index.php
2) The theme of data (e.g., education, environment, transport etc.) and the purpose of data science product are decided by your individual interest and study/professional background in line with the Assignment 1.

open source data relating to manufacturing domain

3) The data science product programming languages (etc. JavaScript, R, or Python etc.) and development platforms (e.g., GoogleColab, RStudio, DJ3, Tableau, or Unity etc.) for the assignment are also decided by yourself.
4) The project will run as a 2-month individual assignment project with the proof-of-concept prototype and project report released in the last week of the project.

**ASSIGNMENT DELIVERABLES**

Body of the report

**1. Project Technical Report**

In the project technical report, you must produce a critical, referenced discussion covering all three Sections below. A report format should be used with cited references that develop your critical discussion and should be taken from current academic literature, such as peer reviewed journal or international conference papers, module text books, published in the last 5 years.

The report produced must be structured with Introduction and Conclusion sections plus three distinct sections each section covering each particular topic as subsections mentioned below:

**Product Design Section**
**Critically discuss the design of your data science product, including:**

- Data source and theme selection and specification
- Application domain/end user's requirements analysis
- Product functional and non-functional requirements specifications
- Product software architecture design
- Product use case specifications

**Product Development Section**
**Critically discuss the development of your data science product, including:**

- The selection of appropriate software tools/platforms and hardware methodologies (e.g., Desktop, Web, Mobile/BYOD, VR/AR)
- Product development software engineering methodology (e.g., Waterfall, Agile/Rapid Prototyping)
- System testing method
- User evaluation plan and methods

**Project Management Section**
**Critically discuss the project management of your data science product, including:**

- Time management with Gantt Chart
- Risk assessment on personal information protection and data security/governance
- Quality control on software development
- Basic Customer/User relationship management
- Basic Product marketing strategy

Word Limit for the assignment is **2000 words** a word count **MUST** be included on your report.

Your report may be subject to checks for originality, which can include use of an electronic plagiarism detection service. Where you are asked to submit an individual piece of software work, the work must be entirely your own.

**References**

It is expected that you use appropriate academic literature relevant to the above concepts introduced in the module. All data sources used and cited in the report should be correctly referenced utilising the Harvard Reference System. References from commercial and other web pages are to be use very sparingly. Excessive use of such web pages will be reflected in a reduction of the use of references mark. A general reading bibliography **MUST NOT** be included. The total number of references is **10-15**.

## 2. Product Prototype

The product prototype source code and user Readme file **MUST** be zipped and submitted with the report together.

**Further Information for Writing up Your Report**

- Software Engineering Methodologies:

https://en.wikipedia.org/wiki/Software_development_process

https://www.tatvasoft.com/blog/top-12-software-development-methodologies-and-its-advantages-disadvantages

- Product Functional and Non-functional Requirements:

https://reqtest.com/requirements-blog/understanding-the-difference-between-functional-and-non-functional-requirements

- Customer Relationship Management for Start-ups:

https://www.entrepreneur.com/article/298189

- Product Marketing Strategy for Start-ups:

https://www.ventureharbour.com/ultimate-startup-marketing-strategy

**Project Report and Product Prototype Assessment Criteria**

| **Data Science Product Design Section (15%)** | | |
|---|---|---|
| | **Poor Excellent** | |
| A simple descriptive account, no critical discussion | □ □ □ □ □ | Considerable critical discussion |
| Irrelevant, unrelated discussion and/or material | □ □ □ □ □ | Accurate, pertinent discussion with reference to the product design |
| | | **Mark /15** |
| **Data Science Product Development Section (15%)** | | |
| A simple descriptive account, no critical discussion | □ □ □ □ □ | Considerable critical discussion |
| Irrelevant, unrelated discussion and/or material | □ □ □ □ □ | Accurate, pertinent discussion with reference to the product development |
| | | **Mark /15** |
| **Data Science Product Project Management Section (15%)** | | |
| A simple descriptive account, no critical discussion | □ □ □ □ □ | Considerable critical discussion |
| Irrelevant, unrelated discussion and/or material | □ □ □ □ □ | Accurate, pertinent discussion with reference to the project management |
| Mark | | **Mark /15** |
| **Overall Quality of the Data Science Product Prototype (45%)** | | |
| Poor coding style | □ □ □ □ □ | Excellent coding practice |
| Unorganised packages structure | □ □ □ □ □ | Clear and well organised package structure |
| Cannot install or run | □ □ □ □ □ | Easy to use with excellent User Friendly Interfaces |
| | | **Mark /45** |
| **Presentation and Quality of Report and Printouts (10%)** | | |
| Illogical and without structure | □ □ □ □ □ | Logical and well formatted structure |
| Very poor presentation and quality | □ □ □ □ □ | Excellent presentation and quality |
| No introduction | □ □ | No conclusion |
| **Comment** | | **Mark /10** |
| | **FINAL MARK** | **/100** |