**Programme Title:** Higher Diploma Data Analytics for Business

**Cohort:**  Full Time Sep2023

**Module Title(s):**  Strategic Thinking

**Assignment Type:** Group / Practical Weighting(s): 20 %

**Assignment Title:** CA 1 – Capstone Project Proposal

**Lecturer(s):** James Garza (james@cct.ie)

**Submission Deadline Date:** 29th October 2023 23:59

**Late Submission Penalty:** Late submissions will be accepted up to 5 calendar days after the deadline. All late submissions are subject to a penalty of 10% of the mark awarded.

Submissions received more than 5 calendar days after the deadline above will not be accepted, and a mark of 0% will be awarded.

Method of Submission: Moodle

**Instructions for Submission:** Your work must be uploaded to Moodle.

• 2 or 3-member group: Capstone Project Proposal in Word or PDF format ONLY. The word count: ~1,000. Individual Reflective Report in Word or PDF format ONLY. The word count is ~500.

• **Individual:** Capstone Project Proposal in Word or PDF format ONLY. The word count: ~1,000. No individual Reflective Report

**Feedback Method:** Results posted in Moodle gradebook

**Feedback Date:**

**Learning Outcomes:**

Please note that this is not the assessment task. The task to be completed is detailed on the next page.

This CA will assess student attainment of the following minimum intended learning outcomes:

1. Critically evaluate the relationship between information technology infrastructure and organisational competitive advantage.
2. Critically analyse and select open source and proprietary software with a view to developing IT solutions for business and business-related IT problems.
3. Utilise tools of strategic business analysis to evaluate the current macro and micro business environment with a view to formulating future action plans.
4. Research emerging technologies and critically evaluate their impact on business and business information systems in general

5. Understand the relationship between data gathering/utilisation and business intelligence and its impact on industry policy.

Attainment of the learning outcomes is the minimum requirement to achieve a Pass mark (40%). Higher marks are awarded where there is evidence of achievement beyond this, in accordance with QQI Assessment and Standards, Revised 2013, and summarised in the following table:

Assessment Task: Capstone Project Proposal

**Title:** Select a title for your capstone project.

**Introduction**: Provide a brief overview of the project and its significance. Explain why the capstone project is important, relevant, and interesting. Mention any real-world problems or challenges that the capstone project aims to address. The capstone project should aim to develop possible solutions.

**Objectives:** Outline the specific goals of your capstone project. What do you intend to achieve through this capstone project? Ensure your objectives are clear, concise, and aligned with the project's goals. You should have between three and five objectives. There should be a business objective or hypothesis (not a statistical test hypothesis) that looks to be explored.

**Problem Definition:** Clearly state the problem or challenge the capstone project seeks to solve. Discuss the context of the problem, its impact, and why it's essential to address it.

**Scope:** There should be plenty of scope for the two-semester capstone project. Define the scope of the capstone project. What will be included and excluded from the capstone project, and who will do what? List the boundaries of the capstone project to avoid any ambiguity. Since this is a two-semester project, ample scope should be provided for in-depth analysis and exploration. Describe the planned methods, techniques, and approaches you plan to accomplish in the capstone project. What do you expect to deliver by the end of semester two? Provide a high-level timeline for the capstone project. Break down the project into phases or milestones and estimate the time required for each.

**Data Sources**: Where will you get your data, and how much do you need? Specify the data sources you plan to use in the capstone project. Show evidence of any permissions to access the data and the use of the data.

**Ethical Considerations:** Discuss any ethical considerations associated with the capstone project, especially if it involves sensitive data, user privacy, or potential societal impacts. This will include dataset permissions and use of data permissions. There can be no medical capstone projects.

**Sample Ideas for Capstone Projects:**

Predictive Analytics for Customer Churn in E-commerce.

Increasing company sales

Financial Fraud Detection (Anomaly detection)

Personalised pricing.

Event analysis

Sports predictions (games, players, plays)

Remember that these project ideas are just starting points. Ensure that your project proposal outlines a comprehensive and in-depth analysis that will occur over two semesters.

**Individual Reflective Report:**

• Regarding the individual reflective report, what were your contributions to all aspects of the project, from selecting the dataset to the final presentation and report? Finally, what was your contribution to the creation of the code and report?

• In terms of team dynamics, how did you work well within the team and were able to leverage each other's strengths to deliver a high-quality project? How did you contribute to the communication and were able to communicate effectively, share ideas and provide constructive feedback to improve the quality of the group work? However, if there were some challenges in coordinating schedules and balancing workload, how were you flexible and adapting to changing circumstances?

• In terms of the learning journey, how did this project help you develop a deeper understanding of data analytics / AI?

• Overall, how did this project help you develop valuable technical and collaboration skills that can be applied in future projects?

• Minimum word count of 500.

Further details of the assessment:

a) Groups of two will be self-selected and put in Moodle. No one can leave a group or be kicked out of a group. If there is an issue in the group, all group members will need to meet with the programme lead or dean of CCT. Your lecturer cannot approve group changes. The maximum group size is 2.

b) Support your analysis with references and properly reference ALL sources that you have used. WARNING – If you do not support your work, you will not receive a high mark!

c) WORD COUNT: 1,000 words. You may lose up to 10% of marks if your report is too short or long!

**Submission Requirements**

All assessment submissions must meet the minimum requirements listed below. Failure to do so may have implications for the mark awarded.

• Submissions must be completed by 29th October 23:59.

• Capstone Project proposal report in PDF or Word format of about 1,000 words.

• Individual Reflective Report in PDF or Word format of about 500 words.

• You must submit your work by the deadline date specified. Otherwise, you will be subject to late submission penalties.

• Use Harvard Referencing when citing third-party material.

• Make sure you submit your own work. No plagiarism will be tolerated.

• Include the CCT assessment cover page.

## 

## **Idea Generation**

* **Credit Card Fraud Detection:** <https://www.kaggle.comdatasets/nelgiriyewithana/credit-card-fraud-detection-dataset-2023>
* Potential Project: Develop a fraud detection system using machine learning algorithms that can identify potentially fraudulent activities. Explore the characteristics of transactions that are most often associated with fraud, and enhance the predictive model accordingly.
* **Fraud** **Detection** – Analyse digital transactions, as well as customer

purchase records, and use the data to identify if the transactions are

potentially fraudulent. When the customer uses a digital medium to

make a payment, you can use the generated data with the trained

model to flag the transaction as potentially fraudulent, which can be

later dealt with and reviewed.

<https://www.kaggle.com/code/rhlchd/capstone-project-credit-card->

[Fraud-detection](https://www.kaggle.com/code/rhlchd/capstone-project-credit-card-)

**Credit Card Fraud Detection**

Credit card fraud, a pressing global issue, saw over 271,000 incidents reported in the United States alone in 2020 according to the Federal Trade Commission.

Dataset: <https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud/>

Objectives

**Develop a Robust Predictive Model:** Create a model that accurately identifies fraudulent transactions, prioritising a balance between sensitivity and specificity.

**Feature Analysis:** Identify and scrutinise the features most indicative of fraudulent activities, enhancing model efficacy.

**Optimization for Imbalanced Data:** Implement techniques to navigate through the challenges posed by the imbalanced nature of fraud data, ensuring accurate predictions.

**Model Evaluation:** Thoroughly evaluate the model utilising various metrics, ensuring its robustness and reliability.

**Risk Mitigation Strategies:** Based on model insights, explore strategies that could mitigate risks associated with fraudulent transactions.

**Problem:** The persistent and pervasive occurrence of fraudulent activities in credit card transactions.

**Context:** Amidst the ubiquity of digital transactions, fraudulent activities have proliferated, necessitating robust detection mechanisms.

**Impact:** Financial losses for individuals and institutions, erosion of trust in digital transaction mechanisms, and potential legal implications for financial entities.

**Title:** Fraud Detection: An Analysis of Fraudulent Activities in Online Payments and Machine Learning-Based Fraud Detection Tool (worth 5 points)

**Introduction:**

*(worth 10 points)*

Should cover:

Background (3 points) 50 words

Significance (7 points) 50 words

Stats

Global losses from credit card fraud reached $ 32.4 billion in 2021 and projections are estimated to reach 43 billion in by 2026.

(Source: <https://merchantcostconsulting.com/lower-credit-card-processing-fees/credit-card-fraud-statistics/> ) This stark rise in the rates of credit card fraud is strongly correlated to the increase in use of online banking, shopping and electronic payment methods being the preferred method of payment for millennials and was even further accentuated by the pandemic where people were discouraged to handle and exchange physical currency. In Ireland, 95% of fraudulent payment transactions

Involved card fraud, costing the people of Ireland and its institutions e33.4million in H2, 2022 (Source: <https://bpfi.ie/publications/fraudsmart-payment-fraud-report-h2-2022/> ) 1in 2 adults in Ireland have received fraudulent text messages, attempting to phish their banking data in the previous 12 months.

.

**Objectives:**

*(worth 20 points)* 100 words

Should cover:

**Primary Objectives**

**Machine Learning Model**

Our primary objectives for this project are as follows. By looking at a large dataset available online, we will create a machine learning model that can analyse historical fraudulent data transactions and detect future fraudulent transactions at a high rate of success. This is our end result, but in order to get to this point we have several primary objectives to achieve:

* Data pre-processing
* Model Training
* Model Evaluation ensuring a high probability rate

**Secondary Objectives**

Hypothesis

To build a fraud detection tool aiming to reduce the number of successful fraudulent credit card transactions, resulting in a reduction in criminal transactions, increasing customer trust, security and a higher insurance of integrity of the financial system.

**Problem Definition**

*(worth 15 points)* 120 words

Should cover:

Problem Statement

Context

Impact

**Scope of the Project:**

*(worth 20 points)* 150 words

Should cover:

Inclusions and Exclusions

Methodology

Milestones

**Data Sources**

*(worth 15 points)* 130 words

Data and Ethical Considerations

* Data sources: clear identification of data source(s) and evidence of permission for data use. Also, consider the quantity of data required.
* Ethical considerations: a thorough consideration of ethical aspects related to the capstone project. This should include awareness of any privacy concerns, bias and fairness.

Harvard referencing and in-text citations have been done correctly for all aspects of the proposal.

Should cover:

Dats Requirements,

data acquisitions

data permissions

**Ethical Considerations**

*(worth 10 points)* 120 words

Should cover:

Data use rights, privacy, and impact

Data Use Rights

Our data has been sourced from Kaggle.com. (Source: [link](https://www.kaggle.com/datasets/dhanushnarayananr/credit-card-fraud)) The data is in accordance with GDPR guidelines and is fully anonymised. The license of this data set is “public domain”, covered under CC0.1 Universal which states “You can copy, modify, distribute and perform the work, even for commercial purposes, all without asking permission.” (Source: <https://creativecommons.org/publicdomain/zero/1.0/>)

The data consists of eight features and one million entries. It is in a numerical format (float64). The file size is 61MB, and contains no missing information. The features included are as follows:

(distance\_from\_home,

distance\_from\_last\_transaction,

ratio to median purchase price,

repeat retailer,

used chip,

used pin number,

online order,

fraud

See link for full EDA analysis: <http://localhost:8892/notebooks/Downloads/archive%20(2)/Credit%20Card%20Fraud%20Detection%20Dataset%20EDA.ipynb>

**Privacy**

The data is fully compliant and contains no PII (personally identifiable information), location information, names, phone numbers, social security information, bank details, or addresses. (Source: <https://www.gdpreu.org/the-regulation/key-concepts/personal-data/#:~:text=Examples%20include%20name%2C%20phone%20number,personally%20identifiable%20information%20or%20PII>.)

The impact of analysing and creating an ML model with this data will be to block transactions with a high likelihood of fraud from the source, allowing global card payment transactions to be safer, saving financial institutions millions in administration costs surrounding fraud, not to mention the money saved from blocking the transactions.

**Preliminary Literary Review**

*(worth 15 points)* 120 words

Should cover:

Existing solutions, GAP Analysis and Relevance

**Proposed Methodology**

*(worth 10 points)* 120 words

Should cover:

Approach and Justification

**Expected Outcomes**

*(worth 5 points)* 80 words

Should cover:

**Conclusion**

*(worth 5 points)* 70 words

Should cover:

Key findings and