

**National College of Ireland**

**Project Submission Sheet**

|  |  |  |  |
| --- | --- | --- | --- |
| **Student Name:** | Anujin Ariunbold | | |
| **Student ID:** | X23171286 | | |
| **Programme:** | BSc in Honours in Data Science | **Year:** | 2 |
| **Module:** | Data Mining and Machine Learning | | |
| **Lecturer:** | Arghir-Nicolae Moldovan | | |
| **Submission Due Date:** | 25th of May | | |
| **Project Title:** | Predicting Diabetes Clinical Dataset | | |
| **Word Count:** | 4586 words | | |

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the references section. Students are encouraged to use the Harvard Referencing Standard supplied by the Library. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action. Students may be required to undergo a viva (oral examination) if there is suspicion about the validity of their submitted work.

|  |  |
| --- | --- |
| **Signature:** | Anujin Ariunbold |
| **Date:** | 25th of March |

**PLEASE READ THE FOLLOWING INSTRUCTIONS:**

1. Please attach a completed copy of this sheet to each project (including multiple copies).

2. Projects should be submitted to your Programme Coordinator.

3. **You must ensure that you retain a HARD COPY of ALL projects**, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer. Please do not bind projects or place in covers unless specifically requested.

4. You must ensure that all projects are submitted to your Programme Coordinator on or before the required submission date. **Late submissions will incur penalties.**

5. All projects must be submitted and passed in order to successfully complete the year. **Any project/assignment not submitted will be marked as a fail.**

|  |  |
| --- | --- |
| **Office Use Only** | |
| Signature: |  |
| Date: |  |
| Penalty Applied (if applicable): |  |

AI Acknowledgement Supplement

|  |
| --- |
| **Data Mining and Machine Learning** |

# CA 1 Individual Assignment

|  |  |  |
| --- | --- | --- |
| **Your Name/Student Number** | **Course** | **Date** |
| **Anujin Ariunbold x23171286** | BSc in (hons) in Data Science | 25th of March |

This section is a supplement to the main assignment, to be used if AI was used in any capacity in the creation of your assignment; if you have queries about how to do this, please contact your lecturer. For an example of how to fill these sections out, please click [here](https://libguides.ncirl.ie/useofaiinteachingandlearning/studentguide).

# AI Acknowledgment

This section acknowledges the AI tools that were utilized in the process of completing this assignment.

|  |  |  |
| --- | --- | --- |
| **Tool Name** | **Brief Description** | **Link to tool** |
| **Grammarly** | Grammar and spelling check | https://www.grammarly.com/grammarlygo |
| **ChatGPT** | Large language model used to assist in analysis interpretation, report writing, and formatting improvement as well as the coding | https://chat.openai.com/ |

# Description of AI Usage

This section provides a more detailed description of how the AI tools were used in the assignment. It includes information about the prompts given to the AI tool, the responses received, and how these responses were utilized or modified in the assignment. **One table should be used for each tool used**.

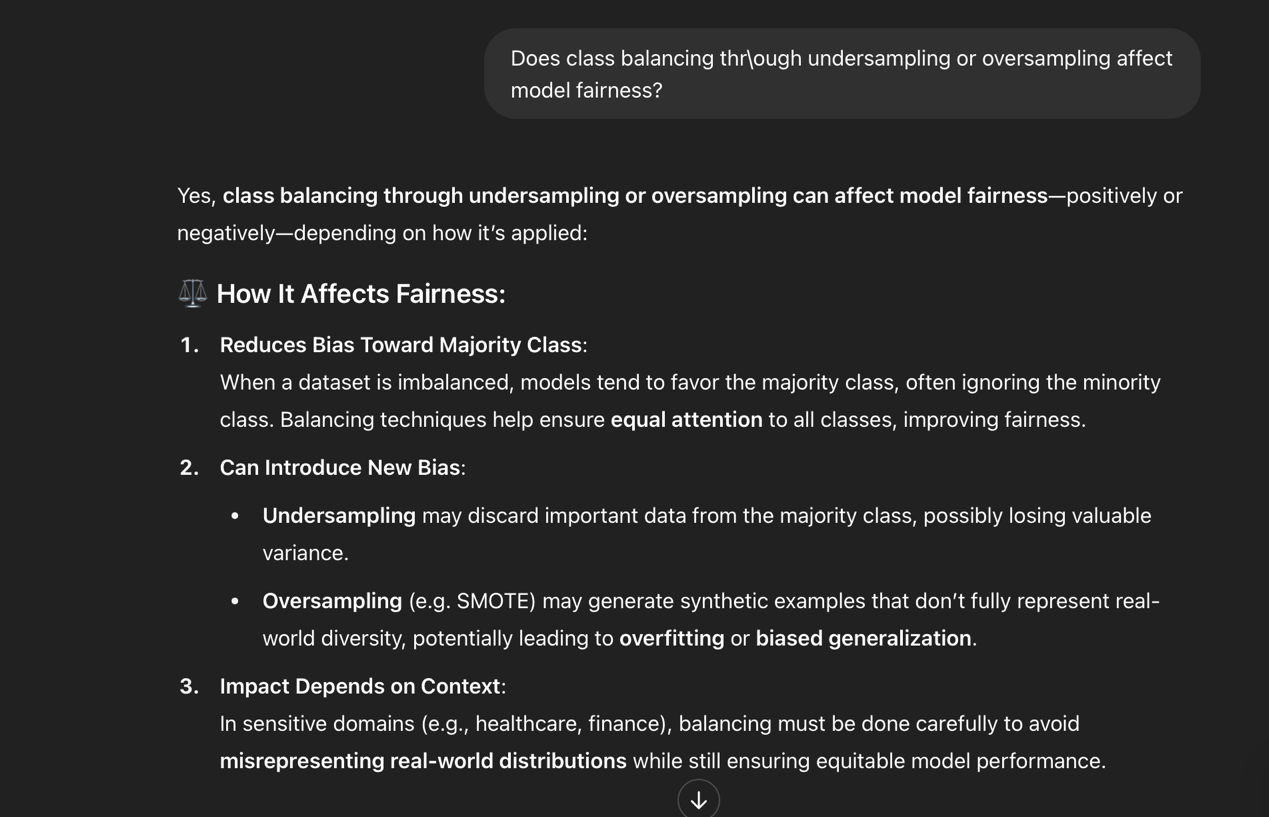
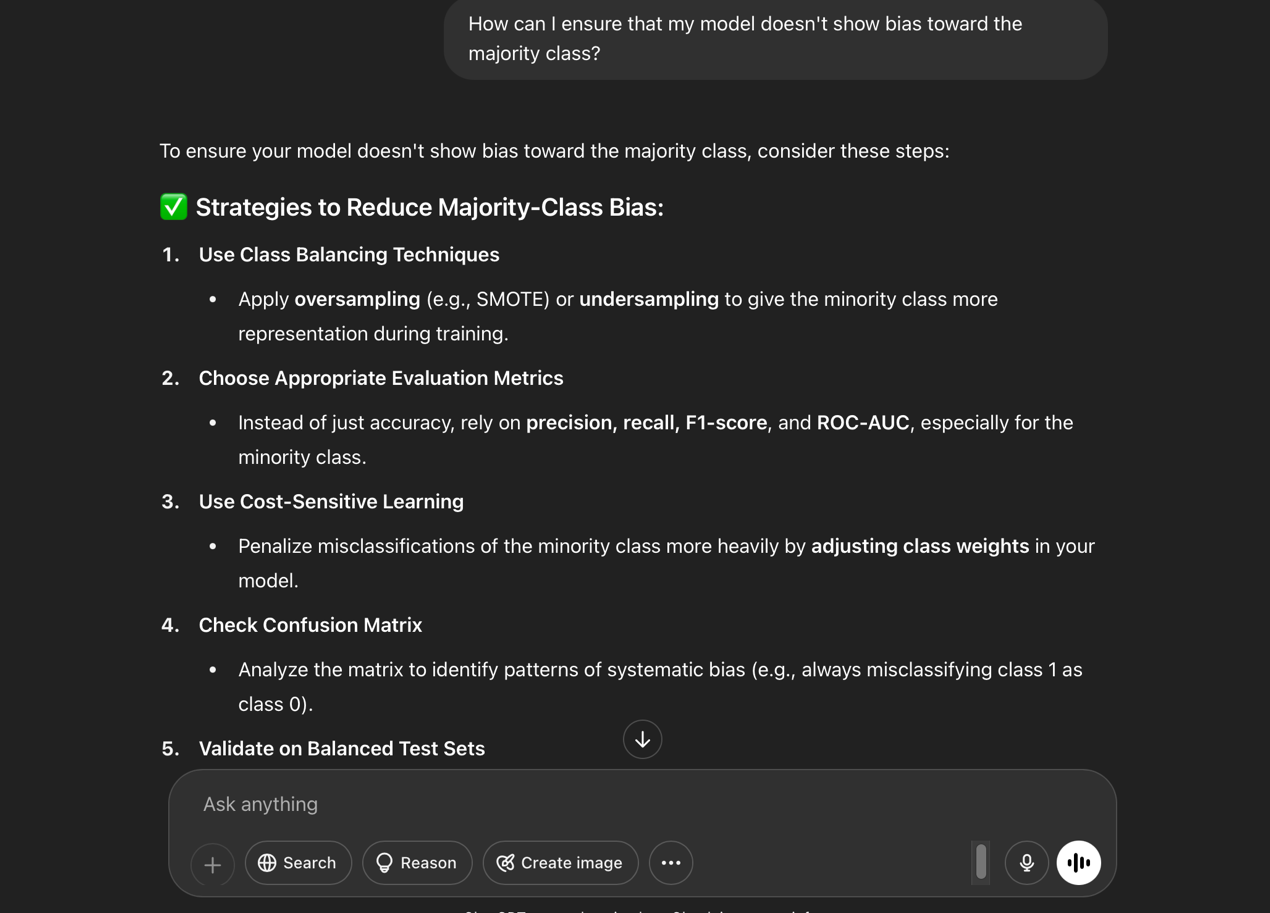
|  |  |
| --- | --- |
| **ChatGPT** | |
| ChatGPT was used to interpret analysis results, write summary reports for each section, improve sentence structure, and assist in code explanation. | |
| Is it necessary to document preprocessing steps for reproducibility and ethical transparency? | Yes, it is absolutely necessary to document preprocessing steps for both reproducibility and ethical transparency |
| Does class balancing through undersampling or oversampling affect model fairness? | Reduces Bias Toward Majority Class: When a dataset is imbalanced, models tend to favor the majority class, often ignoring the minority class. Balancing techniques help ensure equal attention to all classes, improving fairness. |
| How can I ensure that my model doesn't show bias toward the majority class? | trategies to Reduce Majority-Class Bias:  Use Class Balancing Techniques  Apply oversampling (e.g., SMOTE) or undersampling to give the minority class more representation during training.  Choose Appropriate Evaluation Metrics  Instead of just accuracy, rely on precision, recall, F1-score, and ROC-AUC, especially for the minority class. |

|  |  |
| --- | --- |
| Grammarly Go | |
| Grammarly Go was used throughout the writing process for checking grammar, punctuation, and readability of the text. | |
| NA | NA |

# Evidence of AI Usage

A screenshot of a black and white screen

AI-generated content may be incorrect.



# Additional Evidence:

A screenshot of a chat

AI-generated content may be incorrect.

# Additional Evidence: