

Individual Hands-on Assignment – Using Python or R

In this hands-on assignment you are required to conduct AI-driven business analytics of structured and unstructured data using Python or R and submit a report on Moodle course site through Turnitin. The due date of this assignment is on Week 6, Monday 1st of July 2024 at 5:00pm (AEST).

Please note that this assignment is worth 30% of your overall course mark.

1. Requirements

GlobalStore Inc., a leading international online retailer, faces a significant challenge with returned merchandise. The company experienced a substantial volume of their sales returned by unhappy customers, resulting in financial losses and a decline in customer satisfaction. Furthermore, many returned items were damaged and often went to waste, raising environmental concerns, particularly regarding e-waste. Manual handling of e-waste is labour-intensive and prone to errors, leading to increased costs.

To mitigate these issues, GlobalStore plans to harness the power of AI-driven business analytics to accurately predict return likelihood. By identifying potential returns before they happen, the company can take preventive measures to reduce the number of returned products, thereby increasing revenues. Additionally, they aim to employ AI technology to efficiently classify damaged returned items as e-waste. This will not only enhance operational efficiency but also contribute to environmental sustainability and improve customer satisfaction.

In a hypothetical scenario, you are working as a Business Analyst at GlobalStore, and you have been tasked by your manager to conduct AI-driven business analytics (using Python or R). To address the business problem faced by this company, you have decided to:

- Apply Machine Learning to analyse structured data using TWO (2) Machine Learning Algorithms to predict whether an order will likely be returned or not.
- Apply Machine Learning to analyse unstructured data using ONE (1) Machine Learning Algorithm to identify damaged returned items as e-waste.

Please note that the task includes data preparation/cleaning as a pre-processing before conducting the analysis. The dataset required to be analysed includes both structured and unstructured data.

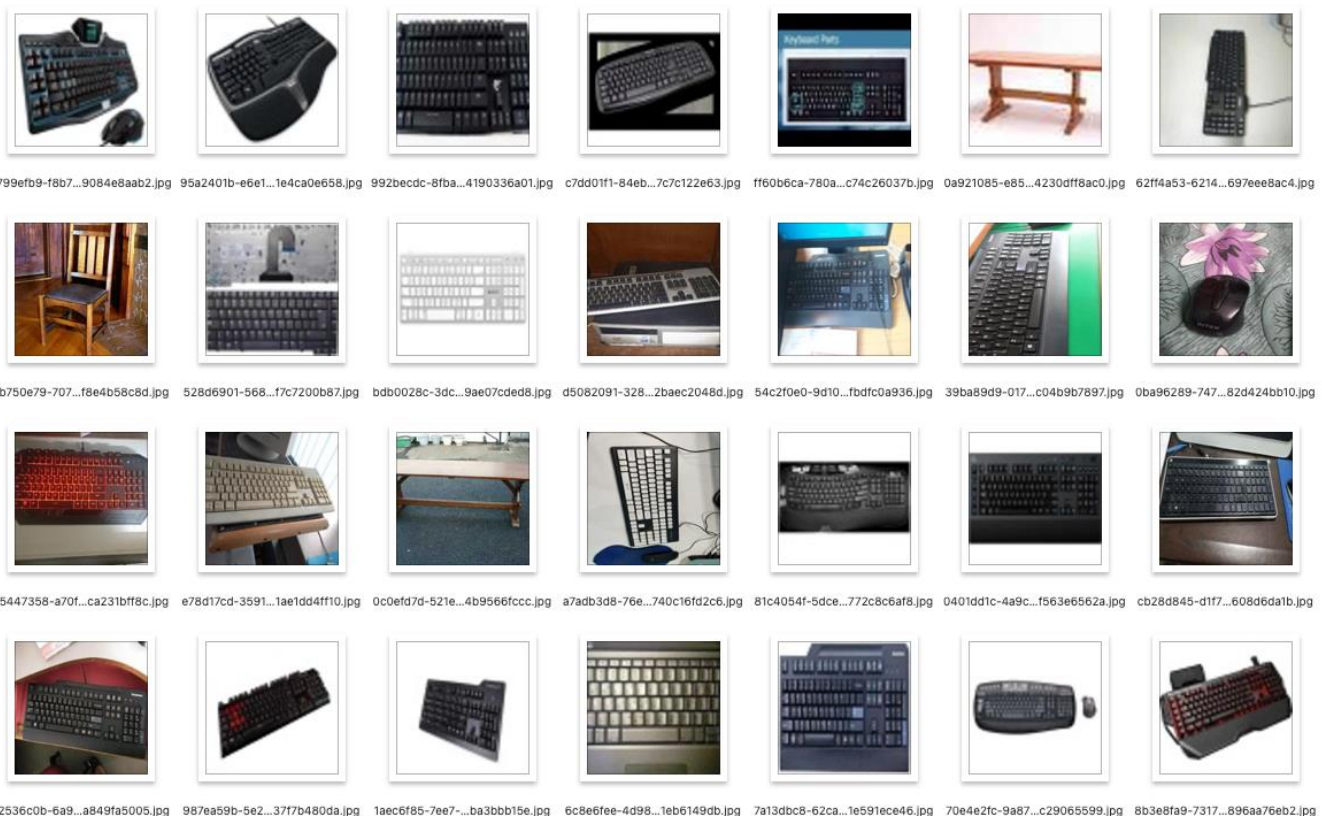
The structured data consists of a large dataset of orders available on Moodle as a CSV file called golabl_store.csv under the Assessments Hub section.

The structured dataset includes the following variables:

- order_id: A unique ID for an order. Multiple rows can have the same order_id, as they correspond to each product in that order.
- order_date: The date the order was placed
- shipping_date: The date the order was shipped

- **shipping_mode:** The shipping method including Standard Class, Second Class, Same Day, and First Class
- **customer_id:** A unique ID for the customer who placed the order
- **segment:** The customer segment including Consumer, Home Office, or Corporate
- **city:** The city to which the order was shipped
- **state:** The state to which the order was shipped
- **country:** The country to which the order was shipped
- **market:** The market to which the order was shipped
- **region:** The region to which the order was shipped including Africa, APAC, Canada, EMEA, EU, LATAM, and US
- **product_id:** A unique ID of the product included in the order
- **category:** The category of product included in the order
- **sub_category:** The sub-category of product included in the order
- **product_name:** The name of product included in the order
- **sales:** The \$ value of the sale of the product
- **quantity:** The quantity of the product ordered
- **discount:** The discount amount applied to the product in the order
- **profit:** The \$ value of the profit made on the product in the order. This can be negative if the Superstore made a loss on the sale of the product
- **shipping_cost:** The \$ value of the shipping cost
- **order_priority:** The priority of handling the order including Low, Medium, High, or Critical
- **returned:** Whether the order was returned or not.

The unstructured data consists of a collection of 2,400 images of returned products such as keyboards, monitors, chairs, tables, etc. This dataset is provided on Moodle as a ZIP file called **product_images.zip** under the Assessments Hub section. Please note that the data is not labelled (not sorted).



1.1. Deliverable

In this assignment you are required to submit a Business Report (in Word format) including the discussion of the findings of your analysis along with justifications and actionable insights.

Your Business Report (3000-word limit) must include the following components:

1. Analysis of Structured Data: predicting item return likelihood

a) Rationale of the choice of ML algorithms (10 marks, 300 words)

Provide justifications of your choice of TWO (2) relevant Machine Learning algorithms to achieve the business objective of GlobalStore to predict whether an order will likely be returned or not.

b) Data pre-processing and selection of the key features (10 marks, 300 words)

Discuss the data pre-processing tasks you decide to undertake along with justifications. How do you handle missing values, outliers, duplicates, etc.

Explain the selection of the most relevant features (key variables) in the dataset required for the ML models, along with justifications.

Discuss the approach you undertake to create the training/testing datasets.

c) Training the TWO (2) ML models (10 marks, 300 words)

Provide an explanation of how each machine learning model is trained. This includes discussing the steps involved, and interpretation of the outputs of the models.

Please include the Python or R code.

d) Testing and evaluating the TWO (2) ML models (15 marks, 350 words)

Provide an explanation of how each machine learning model is tested. Please include the Python or R code.

Evaluate and compare the performance of the TWO (2) ML models in making predictions, along with an explanation of the approach implemented to evaluate their performance.

Provide justifications of the selection of the best performing ML model.

e) Discuss the findings of your analysis and derive actionable insights (15 marks, 450 words)

Discuss the key findings of your analysis along with plausible explanations.

Derive actionable insights from the key findings and translate them into recommendations on how to address the business problem.

2. Analysis of Unstructured Data: detecting e-waste

a) Rationale of the choice of ML algorithm (10 marks, 300 words)

Provide justifications of your choice of ONE (1) relevant Machine Learning algorithm to achieve the business objective of GlobalStore to identify damaged returned items as e-waste.

b) Data transformation (feature extraction) and selection of model parameters (10 marks, 300 words)

loss
optimizer

loss giúp model học tốt hơn
nhớ lại



UNSW
SYDNEY

Discuss the process of feature extraction and the selection of model parameters (e.g. optimal number of clusters). Discuss the findings.

Adam

c) **Implementation of the ML model (10 marks, 300 words)**

Explain the implementation of the Machine Learning model.

Please include the Python or R code

d) **Discuss the key findings of your analysis and derive actionable insights (10 marks, 300 words)**

Discuss the findings of your analysis along with plausible explanations.

Derive actionable insights from the key findings and translate them into recommendations on how to address the business problem.

NB. Throughout the business report, your arguments should be justified and supported with plausible explanations for the key findings and relevant examples.

1.2. Formatting

Word Limit

Each section of the analytics report has a word limit, as indicated in 1.1. Deliverable. The distribution of word count proportionally reflects the complexity and significance of each section, totalling a maximum word length of 3,000 words. There is a (+10%) leeway in word limits for each section.

Please note that tables, figures, diagrams, code and references are excluded from the word count.

You should be mindful of the marks awarded to each section, as indicated in 1.1. Deliverable, when allocating the number of words spend on each section.

Please note that material presented in excess of the word limit for each section will not be considered when grading the assignment.

Formatting

The analytics report should be in 'business report' style (in Word format) with the following requirements:

- Arial 12-point font
- 1.5 spacing
- Page numbers on each page
- Individual Assignment 1 Cover page included (provided on Moodle)
- All required sections included, as indicated in 1.1. Deliverable

Feel free to make whatever use of tables, figures, and diagrams that you believe appropriate and relevant to support your work. Tables, figures, diagrams, code and references do not count towards the word limit.

1.3. Submission

Upload your Business Report document (in Word format) on Moodle.

- You can only upload one report document.



- You are advised to keep a copy of your submission.

The originality of the submission will be checked using Turnitin. Please check the originality report generated by Turnitin during the submission process.

1.4. Late Lodgement & Extensions

Late assignments (without approved extensions) will attract a penalty of 5% of the available marks per day of lateness (including weekends and public holidays). The penalty will be deducted from the mark your assignment is awarded. The assessment will not be accepted after 5 days (120 hours) of the original deadline unless special consideration has been approved. General information on special consideration for undergraduate and postgraduate courses can be found in the [Assessment Implementation Procedure](#) and the [Current Students page](#). Please note:

- Extensions are only granted in exceptional circumstances. You will be required to substantiate your application with appropriate documentary evidence such as medical certificates, accident reports etc.
- You should note that extensions are not usually granted for computer-related problems or heavy workloads (at either your job or University).
- Students are expected to manage their time to meet deadlines and to request extensions as early as possible before the deadline.

1.5. References

If references are cited in your submission, a list of references can be included at the end of the analytics report. The reference list should only list documents / websites that are cited in the assignment (references in text).

Your bibliography and in text citations must be formatted as per the requirements of the Harvard referencing style.

For information on how to acknowledge your sources and reference correctly, see:

<https://student.unsw.edu.au/referencing>

<https://student.unsw.edu.au/harvard-referencing>

1.6. Penalties

Penalties will apply in the following circumstances:

- The assignment contains spelling and grammatical mistakes.
- The submission requirements have not been adhered to.
- Failure to use the Harvard referencing style.
- The assignment is submitted late (5% marks penalty per day of lateness).
- The assignment contains material which is not properly cited in accordance with university policy.
- You should also note that plagiarism or other academic misconduct will not be tolerated, and all instances found will be pursued. At a minimum this typically entails the student being award zero (0) marks the plagiarize assignment.



1.7. Proper academic standards

All assignments are subject to the University's guidelines regarding academic misconduct and as such plagiarism is as unacceptable in this group assignment as it is in other assignment. If plagiarism is found in your assignment it will be fully pursued as per the University (see <https://student.unsw.edu.au/plagiarism> for details).

1.8. Viva Voce

Any student may be called upon to provide a viva voce (from the Latin meaning 'living voice') for any assignment. A viva voce is an interview style meeting where you will be asked to explain, discuss, or use information related to any assignment or work produced for this course. These can be used to ascertain knowledge and ability including the extent to which the student has undertaken the required reading, done preparatory work and can demonstrate understanding of what they have written or presented. Viva voces are used in conjunction with submitted assessment work not instead of submitted work.

This notice is used with permission created by Assoc Prof Lynn Gribble, UNSW Sydney

1.9. Copyright



There are some file-sharing websites that specialise in buying and selling academic work to and from university students.

If you upload your original work to these websites, and if another student downloads and presents it as their own either wholly or partially, you might be found guilty of collusion — even years after graduation.

These file-sharing websites may also accept purchase of course materials, such as copies of seminar slides and handouts. By law, the copyright on course materials, developed by UNSW staff in the course of their employment, belongs to UNSW. It constitutes copyright infringement, if not academic misconduct, to trade these materials.