## Assessment type (🗹):

Questioning (Oral/Written)

Practical Demonstration

3rd Party Report

Other – Project/Portfolio (*please specify)*

**Assessment Resources:**

|  |
| --- |
| Necessary Resources:  * Kaggle or Huggingface account. * A reasonably performant home workstation for out-of-class work. * Google account for Colab (should you wish to use it to secure your model or for datasets). |

## Assessment Instructions:

|  |
| --- |
| Objective: The objective of this assessment is to apply machine learning technologies to automate a task within a workplace environment. You will leverage tools to tackle a real-world or simulated scenario that aligns with your interests or career goals, simulating the capabilities of supervised deep learning models.  In this assessment, you will apply a supervised learning technique to automate a task within a workplace environment. You have the flexibility to select a real-world or simulated scenario that aligns with your interests or career goals. Your goal is to harness the capabilities of supervised deep learning models to accomplish a task that is traditionally done manually.  In order to complete this assessment satisfactorily, you will:  Identify a repetitive, well-defined task from a workplace environment, suitable for automation using a supervised learning technique.  You may select from the list below or come up with your own task:  Categorizing Customer Support Inquiries  Chatbot Conversation Handling  Email Campaign Optimization  Trend Analysis/Predictive Data Modelling  Document Processing Instructions: You should complete the following stages as part of your final project:  1. Business Understanding  2. Data Understanding  3. Data Preparation  4. Choose a Supervised Learning Technique  5. Model Development and Training  6. Model Evaluation and Task Automation  7. Documentation and Reflection  8. Secure and Deploy Your Best Model.  9. Submission Example Scenario for Assessment: Imagine you work for a mid-sized e-commerce company specializing in environmentally friendly products. The company has experienced significant customer growth and as a result, the customer service department has become overwhelmed with the volume of inquiries received daily via email, social media, and the company website.  The company's leadership is interested in technologies that could improve response times and productivity while maintaining a high level of customer satisfaction. They wish to explore AI, ML, and DL solutions to automate parts of the customer service process.  (optional: In discussion with your lecturer, you are encouraged to define your own real-world scenario in your workplace or otherwise of some personal interest to you) Submission Evidence: As part of your submission, you must include:  A copy of your .ipynb file.  A white paper summarizing the project and your findings.  This document containing a link to your model (and notebook) on Kaggle, and a completed process guide deemed satisfactory by your lecturer. |

**Assessment Instrument:**

|  |
| --- |
| AT4: Apply Machine Learning to Task Automation1. Business Understanding: 1. Use the CRISP-DM methodology to analyse the business needs and establish an initial ML work brief.  2. Define the scope of the task, the expected outcomes, and the performance criteria for the model's success.  3. Confirm your work brief and tasks with supervising personnel and with relevant organizational policies and procedures. 2. Data Understanding: 1. Apply stage 2 of the CRISP-DM data process to satisfy the data requirements of your ML work brief.  2. Gather or generate an initial dataset that can be used to train and test your machine learning model.  3. Identify issues in the data and explore the suitability of the initial dataset for your chosen work brief.  *In conducting the above steps, you must:*  Consider the availability and applicability of structured, unstructured, labelled and unlabelled data sources for your work brief.  Confirm whether the data source(s) will satisfy the objectives of the work brief by performing initial exploratory data analysis visualizations, as needed.  Confirm that data attribute names contain relevant target according to work brief.  Identify any necessary data transformations to be carried out on the initial dataset.  Randomise, deduplicate and check your initial dataset for imbalances and biases.  Identify important considerations based on biased or unbiased nature of your dataset. 3. Data Preparation 1. Apply stage 3 of the CRISP-DM data process to arrange and verify your final dataset according to the established requirements your ML work brief.  2. Review the applicability of clustering or other classification algorithms to generate labels for any unlabelled data in your dataset in the context of your work brief.  3. Perform necessary pre-processing steps on the core dataset, including:  Data cleaning,  normalization,  grouping of labelled and unlabelled data,  application of labelling techniques on unlabelled data,  one-hot encoding for any categorical variables,  splitting final dataset into training, validation, and test sets.  4. Produce a final dataset relevant to the task chosen that includes appropriately labelled data suitable for supervised training.  *In conducting the above steps, you must:*  Divide data into training subset and evaluation subset according to work brief  Confirm that data is correctly grouped as labelled or unlabelled  Confirm that default and non-default training parameters control required learning algorithm according to work brief  Use selected parameter and feature engineering on required training data 4. Choose a Supervised Learning Technique: 1. Research and analyse the applicability of a range of supervised learning techniques, considering your final dataset and the key objectives of your work brief.  2. Choose at least one supervised learning technique (e.g., decision trees, support vector machines, Naïve Bayes, neural networks (transformers, CNN, RNN)) to apply to task automation.  *Note: Some factors you may wish to consider include task relevance, data suitability, expected performance, interpretability, and computational efficiency.*  *In conducting the above steps, you must:*  Analyse regression algorithms, decision trees or neural net algorithms for labelled data, where required  Analyse clustering, association, instance-based or neural network algorithms for unlabelled data, where required  Select algorithm for dataset according to analysis findings   5. Model Development and Training: 1. Confirm expected ML outputs with required personnel.  2. Set up your coding environment using Google Colab, Kaggle Kernels, or a local Jupyter notebook setup.  3. Define and load your chosen model. If needed, define the model architecture and size.  4. Conduct training and validation using your selected supervised learning technique.  5. Compare expected and actual outputs. If applicable, apply relevant loss function and back propagation techniques.  6. Review training process and make any necessary refinements.  *In conducting the above steps, you must:*  Confirm expected ML outputs with required personnel  Compare expected and final outputs with required personnel  Set machine training data parameters according to work brief  Select model size according to work brief  Set test data parameters according to work brief  Use selected parameter and feature engineering on required test data  Identify and rectify any functionality issues in test dataset  Finalise test data procedures according to work brief   6. Model Evaluation and Task Automation: 1. Evaluate target data outputs against initial performance criteria.  2. Evaluate the effectiveness of your model using the unseen (by the model) test dataset.  3. Select the most relevant metrics (e.g., accuracy, F1-score, precision, recall) for your particular use-case. Document relevant metrics and accuracy.  4. Demonstrate the effectiveness of the model in automating the chosen task by running organizational data through the model. Determine whether it performs as expected or not. If not, hypothesize some reasons why it may not be performing as intended.  *In conducting the above steps, you must:*  Document analysis findings according to organisational policies and procedures  Run variables through selected algorithm according to work brief  Compare expected and actual ML outputs  Adjust algorithm and re-run variables through selected algorithm according to work brief  Confirm that new algorithm outputs yield accurate output results  Compare expected and final outputs with required personnel  Finalise machine training data procedures according to work brief  Set validation data parameters according to work brief  Use selected parameter and feature engineering on required validation data  Identify any functionality issues of parameters  Refine ML parameters according to work brief  Review target data outputs according to work brief  Adjust model based on any discrepancies of outputs, where required  Run variables through ML model and record outputs  Compare outputs returned by ML model against target data outputs   7. Documentation and Reflection: 1. As you complete the process of training your model, you should clearly document and comment on each step in your notebook, including data processing, model architecture decisions, training processes, and evaluation results.  2. Discuss any potential ethical considerations, biases in the dataset, limitations you faced, and future improvements that could be made in relevant sections of your notebook.  3. Summarize and expand on your IPython notebook in a white paper using the white paper template as a guide.  Remember: this white paper has a different audience from your technical notebook.  Your white paper will serve as a summary and documentation of your work for others in your company to reference and understand what you have done, even if they lack a technical background.  At a minimum it should contain:  A summary of your data preparation methodologies and rationale.  A justification of why your algorithm selection aligned with the project brief.  A summary of your model training and evaluation processes and findings.  A discussion of ethical considerations, data governance, and legislative requirements in line with Australia's AI Ethics Framework.  A reflection on the adaptability of the ML model to real organizational contexts and any challenges it may face in the real world.  *In conducting the above steps, you must:*  Run organisational data through algorithm according to work brief  Record predictive accuracy of ML model according to work brief  Document metrics and accuracy of ML data predictions according to organisational policies and procedures 8. Secure & Deploy Your Best Model: 1. For the purposes of this assignment your organization seeks to make AI accessible for all, as a result it is their policy to store all the results of their research into AI in a publicly accessible repository (Kaggle or Huggingface).  2. As a result of your work, publish your best performing model to Kaggle or Huggingface and provide a link to your model here:  3. Upload your IPython notebook to Kaggle, make it publicly accessible, and provide a link to it here:  *In conducting the above steps, you must:*  Secure and save ML model according to organisational policies and procedures 9. Submission: 1. Export your Jupyter notebook as an .ipynb file, ensuring it runs without errors from start to finish.  2. Submit the .ipynb file along with your white paper and this document. |

## Observation Checklist

|  |  |  |  |
| --- | --- | --- | --- |
| Checkpoint | Done | Give a brief description | S/NYS |
| Confirm your work brief and tasks with supervising personnel and with any relevant organizational policies and procedures. |  |  |  |
| Confirm input machine training data source according to work brief |  |  |  |
| Confirm that data attribute names contain target according to work brief |  |  |  |
| Confirm that default and non-default training parameters control required learning algorithm according to work brief |  |  |  |
| Confirm that data is correctly grouped as labelled or unlabelled |  |  |  |
| Confirm expected ML outputs with required personnel |  |  |  |
| Confirm that new algorithm outputs yield accurate output results |  |  |  |
| Compare expected and final outputs with required personnel |  |  |  |