Imagine you are a data scientist investigating an interesting problem in your domain, and you are tasked with proposing a machine learning solution.

Steps:

1. Pick a domain of your interest. Then, you may try the following:
   1. Pick a problem which already exists in your domain.
   2. Find a client / stakeholder and understand their domain problem.
2. Define the problem statement.
3. Detail a plan on how to:
   1. Find / collect data (you may collect your own or use any available dataset). Some notable dataset repositories are as follows:
      1. Kaggle (<https://www.kaggle.com/datasets>)
      2. UCI Dataset (<https://archive.ics.uci.edu/ml/index.php>)
      3. Google Dataset (<https://datasetsearch.research.google.com/>)
   2. Clean the data / perform exploratory data analysis (EDA) to get valuable insights, which includes:
      1. Data preprocessing (data cleaning)
      2. Data visualisation
   3. Pick several machine learning methods which you think is suitable to solve your problem.
   4. Perform the training.
   5. Evaluate your model using performance metrics.
4. Pitch your solution to your client / stakeholder (in this case, it's your panel) and get them onboard. Revise your plan if necessary. Then pitch again.
5. Execute your plan.
6. Document everything in your Jupyter / Colaboratory notebook using markdown and figures.

Example notebook:

<https://www.kaggle.com/nadintamer/titanic-survival-predictions-beginner/comments#Titanic-Survival-Predictions-(Beginner)>

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| No. | Criteria | Grading | Weightage | Total Score |
| 1. | **Introduction:**  Description of proposed ML project idea, the AI application and the industry involved. | 1. Did not present correct information - 1 mark 2. Present minimal information - 2 mark 3. Present partial information - 3 mark 4. Present sufficient information and appropriate justification - 4 mark 5. Present excellent information and sound justification supported by reliable sources - 5 mark | 0.5 | 2.5 |
| 2. | **Problem Statement:**  Discussion on the issue/problem to be resolved. | 0.5 | 2.5 |
| 3. | **Dataset:**  Description of the dataset used in the application, and if the dataset was collected by the students, where the dataset was taken or how it was collected.  Discussion and justification of data preprocessing methods using appropriate visualisation, supported by reliable sources. | 1 | 5 |
| 4. | **Proposed Approaches:**  Discussion and justification on ML models/algorithms used and the ML tool used in the ML application, supported by reliable sources. | 1 | 5 |
| 5. | **Conclusion:**  The strength and positive impact of the proposed ML project idea towards industry/society | 0.5 | 2.5 |
| 6. | **Notebook Report:**  Detailed documentation with justification from reliable sources on following items:   * Background of the problem statement * Dataset used and its collection method * Methods of preprocessing justified by insights drawn from data explorations * Justification on selected ML algorithms, methods and evaluation metrics used | 0.5 | 2.5 |
| 6. | **Interpersonal and Communication Skills** | 1. Poor - 1 mark 2. Low - 2 mark 3. Fair - 3 mark 4. Good - 4 mark 5. Excellent - 5 mark | 1 | 5 |
| 7. | **Bonus point 1:**  Hosting the ML notebook on Github | 1. No - 1 mark 2. Yes - 5 mark | 0.5 | 2.5 |
| 8. | **Bonus point 2:**  Contribute a detailed article on Medium / blog / LinkedIn. | 1. No - 1 mark 2. Yes - 5 mark | 0.5 | 2.5 |
| **Grand Total** | | | | **30** |
| **Normalised Total (Grand Total / 3)** | | | | **10** |