CSC2114 – Week Five Instructions
Project Narrowing, Dataset Review, EDA, Model Training, and Conceptual Integration
https://forms.gle/mjTLN82xkoDMLCfT6

Get an approved Project title (I should review and approve your project title)

1. Narrow Down to a Specific Project

- Select and clearly **state the specific project** you are pursuing.
- Frame your **computational problem**:
  - What is the **input** (dataset, features)?
  - What is the **output** (prediction, classification, clustering, etc.)?
  - What is the **objective** (accuracy, efficiency, explainability, etc.)?
- Clearly indicate whether your task is **supervised, unsupervised, or semi-supervised learning** and explain *why* this paradigm fits your problem.

2. Dataset Review and Selection

- Review **2–3 potential datasets**. For each dataset, explain:
  - Source (must not be Kaggle).
  - Variables/features and dataset structure.
  - Suitability for supervised/unsupervised/semi-supervised learning.
  - Strengths and limitations.
- Select **one final dataset** and justify your choice.

Rule: You must use **Kaggle Notebooks** for coding and EDA, but datasets must not come from Kaggle.

3. Exploratory Data Analysis (EDA)

- Conduct a full **EDA in Kaggle Notebooks** on your chosen dataset. Include:
  - Dataset overview (dimensions, data types, structure).
  - Missing values and cleaning approach.
  - Basic descriptive statistics.
  - Visualizations (histograms, scatterplots, correlation heatmaps, distributions).
- Report not just what you see but **what it means** for your problem.
  - Example: "The class imbalance revealed in the target variable means accuracy is not a sufficient metric; F1-score will be used instead."

4. Model Fitting and Evaluation

- Train and evaluate at least **one AI model.**
- For the model you use, explain:
  - Why you chose this model (connection to the problem and dataset).
  - The model's architecture/parameters.
  - The evaluation metric(s) chosen and why.
  - Results and interpretation (performance achieved, limitations noticed).

5. Integration of Core AI Concepts

In your presentation and report, you must explicitly demonstrate how covered concepts are guiding your choices:

- **Learning Paradigm:** State whether your project is supervised, unsupervised, or semi-supervised and why.
- **Types of Intelligence:** Which of Gardner's intelligences best reflects your project? (e.g., linguistic intelligence for NLP, naturalistic intelligence for agricultural datasets).
- **Data Mining:** Which approaches are being applied? (classification, clustering, anomaly detection, etc.).
- **Knowledge Representation & Reasoning:** How would your project represent knowledge? Are you using rules, graphs, or embeddings? What kind of reasoning does your model embody?
- **Philosophical Framing:** Reflect briefly—does your approach lean more towards Weak AI (task-focused) or suggest steps toward Strong AI? How do debates like the Turing Test or the Frame Problem apply?

6. Reporting and Presentation
Here is the template:
https://docs.google.com/presentation/d/1ipDkBmqOLALBvqGlaSBxRiKZid7R9mMO/edit?usp=sharing&ouid=100050837801628017110&rtpof=true&sd=true

- Prepare a **PowerPoint presentation** using the template provided. Your slides must cover:
    1. The specific project and computational problem.
    2. Dataset review and justification of the final dataset.
    3. EDA results (figures, charts, interpretations).
    4. Model description, architecture, evaluation, and results.
    5. Integration of concepts (learning paradigm, intelligences, data mining, KR&R, philosophy).
- Record a **video presentation** of your slides:
    - Duration: **≤15 minutes.**
    - Your **face must be visible.**
    - Present clearly and professionally.

7. Submission : https://forms.gle/mjTLN82xkoDMLCfT6

- Upload your **slides** and **YouTube video link** (public or unlisted) via the Google Form provided.
- Include any supporting dataset description files if requested.

Important Notes : https://forms.gle/mjTLN82xkoDMLCfT6

- This is a group-based submission
- All coding must be in **Kaggle Notebooks.**
- Datasets must **not be from Kaggle.**
- AI-generated content will attract **severe penalties.**
- Use the presentation Template provided