

Data Science Capstone Project: Research-Informed Investigation

Project Objective

You will explore a topic of current research in data science or machine learning, find and read at least five recent peer-reviewed papers or preprints, and either:

1. Replicate and evaluate key findings on multiple datasets, or
2. Summarize the state of the field and propose a small but novel extension or experiment.

You will write a full research-style report in Overleaf (LaTeX) and submit all code, results, and analysis. AI tools (such as ChatGPT, Copilot, or Claude) may assist in code generation, but all results and claims must be verified, explained, and critically analyzed by you.

Project Timeline

Phase 1: Proposal Submission (Tuesday June 24)

Submit a one-page PDF proposal, which includes:

- A working title
- A clear research question explaining your goal, why it is important, why it is timely, etc.
- Brief summaries of 5 research papers
- Description of your planned experiment or synthesis
- Datasets you plan to use (include links if possible)
- What you expect to contribute, replicate, or test

Phase 2: Final Paper + Code + Checklist (Thursday June 26)

Submit your final:

- Overleaf report (IEEE/ACM format)

- Code and results (via GitHub or ZIP)
 - Project checklist (see below)
-

Checklist: Project Deliverables

Your final submission must include:

1. Literature Review
 - a. Five peer-reviewed (preferably IEEE) or arXiv papers (from 2020 onward)
 - b. Concise summary of each paper
 - c. Compare/contrast approaches
 - d. Identify gaps or open questions
2. Experiment / Analysis which includes at least one of the following:
 - a. (Replication): Recreate results or core insights from 1–2 papers using at least three related datasets
 - b. (Synthesis): Combine or extend ideas in the literature and test a small hypothesis (on at least three data sets)

Notes:

1. Code may be aided by LLMs but must be:
 - a. Run and verified
 - b. Clearly explained
 - c. Include error handling, plots, and interpretation
2. Final Report in Overleaf should preferably include some or all of the following:
 - a. Title, abstract, introduction
 - b. Related work
 - c. Methodology
 - d. Experimental setup and datasets
 - e. Results (with tables, figures)
 - f. Conclusion and future work
 - g. References (BibTeX, IEEE/ACM format)