# CS 520
# Final project report

Reminder of upcoming final project deadlines:

- You will present your final project to your TA in a group meeting **between Thursday, December 10 and Tuesday, December 15, 2026**.
- The final project report is due **Tuesday, December 15, 2026, 11:55 PM EST**.

This is the last deliverable of the project. It builds on your project plan (submitted October 8, 2026) and on the feedback you received from your TA at your mid-date report meeting.

Each team submits a single final project. All team members are expected to contribute substantially to the work and to the write-up.

## What to submit

Your final project consists of four (4) parts:

1. **A final presentation** to your TA (described below).
2. **A written report**, formatted as a research paper (described below).
3. **A replication package**: the code, scripts, and data (or instructions for obtaining the data) needed to reproduce every result in your report.
4. **A contributions statement** describing what each team member did.

## The final presentation

As with the mid-date report, instead of presenting in class, your team will have a group meeting with your TA (the TA assigned to your project) between Thursday, December 10 and Tuesday, December 15, 2026. The lectures on Thursday, December 10 and Tuesday, December 15 are cancelled to make time for these meetings. Contact your TA to schedule a meeting time during that week. Do this early, so that you can find a time that works for your TA and for all team members. All team members are expected to attend. A student who misses the meeting without a valid excuse will lose 20 points (out of 100) individually on the final project. You must be ready to present on Thursday, December 10, 2026, even if your meeting is later.

At the meeting, your team will give a 7-minute presentation describing your project and its results, followed by a few minutes of questions and feedback from your TA. Bring slides. You will also submit your slides on Canvas along with your final report by Tuesday, December 15, 2026, 11:55 PM EST (see Deliverables below). The time limit is strict. Your presentation should cover:

1. **The Problem.** Your research question(s), and (**MSR**) the dataset(s) you used or (**REPL**) the paper you replicated and your extension's research question. Put a one-sentence summary on the slide.
2. **The Design.** How you designed your experiments: the data, the processing pipeline, and what you measured.

3. **The Results.** What you found for each research question, and how you know your results make sense. For the **Research paper replication and extension**, how your replicated results compare to those in the paper. This should be the bulk of your presentation.

4. **The Conclusions.** What your results mean, the main threats to their validity, and what you learned.

Use the feedback you receive at this meeting to improve your report before it is due.

## The report

Your report should read like a short software engineering research paper. You may use any tool or template you like to write it, as long as the main text uses an 11-point font and the page has 1-inch margins on all sides on letter-size paper. The 11-point requirement applies only to the main text; figures, tables, captions, footnotes, and references may use other sizes. The report should be at most 10 pages, not including references and appendices. Shorter is fine if it is complete; padding will not help. You may add appendices (e.g., with additional results or tables), but the course staff may not read them when grading the project, so everything needed to understand and evaluate your work must be in the main 10 pages. (The contributions statement appendix, described below, will be read.) The report should contain at least the following sections:

1. **Introduction.** What problem are you studying, why does it matter, and what did you find? State your research questions explicitly and summarize your answers to them. A reader should be able to understand your main findings from the introduction alone.

2. **Background and related work.** Describe the work your project builds on and how your work relates to it. Cite published research papers, not just websites.

3. **Methodology.** Describe your data, your processing pipeline, and your experimental design for each research question, in enough detail that another team could repeat your study. (For the **Research paper replication and extension**, Methodology and Results are organized differently; see below.)

4. **Results.** Answer each research question, using tables and figures where appropriate. Report appropriate statistical tests, if relevant. Make your findings clear: say what you found, not only what you did.

5. **Discussion and threats to validity.** What do your results mean? What could make your conclusions wrong (e.g., biases in the data, measurement error, confounding factors, too few samples), and what did you do to mitigate these threats?

6. **Conclusion.** Summarize your contributions and findings, and identify interesting future work.

In addition, the report must address the following, depending on your project type:

**MSR 2027 Mining Challenge.** State which dataset(s) you used (GitSkills, SpecMine, or both), how you selected, filtered, and processed the data, and how many data points each of your analyses is based on. Explain how your research questions relate to the challenge description (https://2027.msrconf.org/track/msr-2027-mining-challenge).

**Research paper replication and extension.** Instead of single Methodology and Results sections, your report must have two sections, Replication and Extension, each with its own Methodology and Results subsections:

- **Replication.** Which paper did you replicate, and which of its research questions? Describe the process of replicating the experiments, including what you had to change or fix to get the paper's code to run. Compare your results to those reported in the paper, explain any deviations, and discuss whether those deviations are meaningful.
- **Extension.** What new research question did you ask, and why is it interesting? Describe the experimental design you used to answer it, the new data or benchmarks you used, and your results.

Finally, describe lessons learned from the replication: what was easy, what was hard, and what the original authors could have done to make their work easier to replicate.

### The replication package

Submit a link to a Git repository (e.g., on GitHub) containing all the code and scripts you wrote, along with a `README` that explains how to install the dependencies, obtain the data, and run your experiments to reproduce every table and figure in your report. If the data is too large to include in the repository, explain how to obtain it. Make sure the course staff can access the repository (either make it public or give the staff access).

### The contributions statement

Include a contributions statement as an appendix to the report (it does not count toward the page limit). For each team member, list, in a few bullet points, the concrete parts of the project that member was primarily responsible for. For example:

- getting the paper's code running, or building the data extraction pipeline;
- designing and running the experiments for a particular research question;
- the data analysis and statistical tests;
- writing particular sections of the report, or preparing the slides.

Be specific: "worked on the experiments" is not helpful; "wrote the scripts to run the replication on the paper's benchmark and produced Table 2" is. Several members may share responsibility for the same part. All team members must agree on the statement. If your team cannot agree on it, contact the professor before the deadline.

### Deliverables

Meet with your TA between Thursday, December 10 and Tuesday, December 15, 2026 to present your final project (see above).
By **Tuesday, December 15, 2026, 11:55 PM EST**, one member of each team should submit via Canvas:

1. The report, including the contributions statement appendix, as a single .pdf file.
2. A link to the replication package repository.
3. The slides from your final presentation, as a .pdf file.