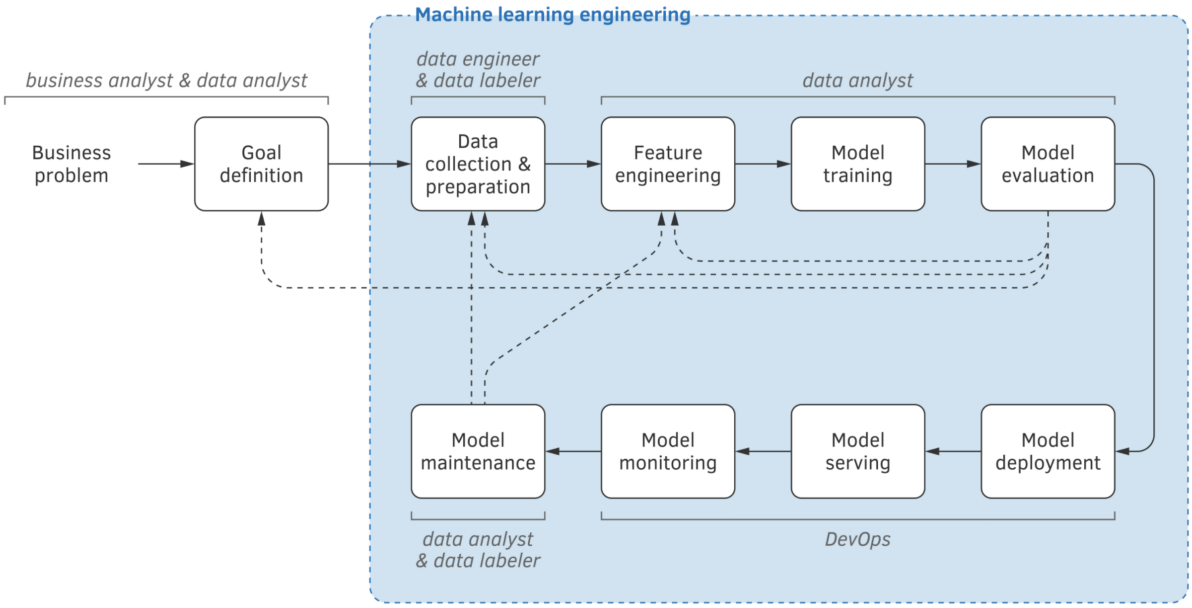
REI603M - Final Project
Spring 2024



Final Project: Building from Scratch

Overview

In this final project, you will present your project focused on designing and developing a machine learning system from the ground up. The goal is to iterate and improve upon your work in the course.

To demonstrate your progress throughout the AI lifecycle, you will deliver a 20-minute presentation that highlights your contributions at each stage. Your final grade will be based on your performance in each part of the lifecycle and your ability to develop a complete AI software solution. **Please structure your presentation according to the stages outlined below.**

Product Development Projects: Stages of the AI Lifecycle

Problem Definition **(5%)**

Begin your presentation by clearly defining the problem you are attempting to solve and explaining its significance.

- Why is this problem important, and what are the consequences of solving or not solving it?
- What are the limitations of current solutions? Remember to cite sources to place your work within the body of work that has already been done.
- What performance metrics should be considered, and why are they relevant?
- If applicable, what key performance indicator (KPI) would be relevant for a business trying to solve this problem?
- Why is machine learning (ML) an appropriate solution for this problem?
- How could ML create value for individuals or businesses attempting to solve this problem?

Data Collection **(15%)**

Describe the data collection process for your project.

- What data did you collect, and how did you collect it?
- If using LLMs, how did you structure your prompts? What experiments did you do to find the best prompt?
- How did you label the data, if applicable?
- Provide basic summary statistics about your dataset, such as size, class distribution, and notable patterns, surprises or outliers.

Data Preparation and Feature Engineering **(15%)**

Explain the data preparation and feature engineering steps you took.

- What data preparation was necessary to prepare your dataset for modeling and/or prompting?
- What features did you use in your model, and did you build any new features?
- Did you eliminate any features from your dataset, and if so, why?
- What data cleaning did you perform to remove or fix bad examples?
- For LLMs, do you extract information in a structured manner? What structure do you impose?

Baseline Models **(10%)**

Discuss the baseline models you used as a starting point for your machine learning solution.

- Describe the baseline models you used and explain why you chose them.
- If relevant, compare your baselines to other benchmarks or industry standards.
- Present more than one baseline, if applicable, and discuss how they compare to one another.

Model Training and Evaluation **(20%)**

Detail the model training and evaluation process.

- Discuss the various models you tried on your data and explain your choices.
- Describe how the models performed on a test set and compare their performance to the baselines.
- If applicable, explain any hyperparameter tuning you performed and its impact on your results.
- If applicable, show a learning curve for your best model built using cross-validation (sample at least 5 points for each amount of training data).

Model Deployment and Monitoring **(15%)**

Demonstrate how you deployed your machine learning model in a cloud service and how you are monitoring its performance.

- Describe the deployment method you used, including the cloud service, specific method, and any challenges encountered.
- Explain how you are monitoring your deployed model, including the tools used, metrics tracked, and any alerts or notifications you have or would set up.

Model Maintenance **(10%)**

Discuss potential improvements to your machine learning model and how you would approach another round of the AI lifecycle.

- Consider ways to improve your model's performance or efficiency, such as experimenting with different algorithms, features, or hyperparameters, or exploring new data sources.
- Discuss the time and cost associated with each potential improvement and prioritize accordingly.
- Consider ways to automatically update your model with new data to ensure it remains accurate and relevant over time.
- Explore ways to collect labeled data from users or other sources to enable continuous learning and improvement after deployment.

Lessons Learned **(5%)**

Reflect on your experience and identify the most important takeaways from your project.

- Highlight insights or learnings that stood out to you throughout the AI lifecycle.
- Identify the most important lesson learned during the project and explain its significance for your future work in the field.

Research Project: Stages of the Research Project

Research Question and Motivation **(10%)**

- Clearly state your research question and explain its significance in the context of current research.
- Discuss the potential impact of answering this research question on the field of study and its applications.

Related Work **(20%)**

- Provide an overview of the existing literature related to your research question.
- Identify gaps in the current knowledge and explain how your research aims to address these gaps.
- Discuss the strengths and limitations of the most relevant related works.

Methodology **(30%)**

- Describe the methods and techniques you employed to answer your research question.
- Explain the rationale behind your choice of methods and discuss any novel contributions or modifications you made.
- Detail any experiments, simulations, or analyses you conducted and the datasets you used.

Results and Discussion **(30%)**

- Present the main findings of your research and interpret them in the context of your research question.
- Discuss the implications of your results for the field of AI and potential applications.
- Address any limitations of your study and propose future research directions.

Draft Manuscript **(10%, graded separately)**

- Prepare a draft manuscript that follows the structure and formatting guidelines of a relevant conference or journal.
- Ensure that your manuscript clearly communicates your research question, methodology, results, and conclusions.

Deliverables

To complete this final project, you must:

- Submit the code for your machine learning project (product development track) or your draft manuscript (research track). Please refrain from submitting datasets and submit this as a `.zip` file.
- Deliver a 20-minute presentation that covers all the key points of your chosen track.
- If working in a group, both members should participate in the presentation.
- Grade each other using the rubric provided.

Your grade for the final project is 60% the grade from the instructor, 30% the grade from peers and 10% for participating in the peer review.

If you have any questions or need clarification, please contact the course instructor at `hafsteinne@hi.is`. When seeking assistance, provide context and explain what you have tried so far to help the instructor understand your thought process and provide more targeted guidance.

The deadline for submitting your project is right before your assigned presentation in either of the last two weeks of the semester. Manuscript drafts can be submitted until the end of April.