

# MA289 Intro to Statistical Learning

## Project 3: Applied Demonstration of an AI Modeling Method

MAJ Patrick Kuiper

**DUE No Later Than 14 April 2026 at 23:59**

Submit:

- A **PDF report** in Canvas.
- A public **GitHub repository** containing clean, organized code.
- A **5–8 minute teaching presentation** including live coding.

### Overview

This project builds directly on Assignment 2. You will extend your original pathological synthetic demonstration by incorporating a **real-world dataset** that exhibits similar structural challenges or properties. The objective is to demonstrate the power, limitations, or necessity of your selected modeling method in both controlled (synthetic) and real-world contexts.

Your project must:

- Refine and improve your synthetic experiment.
- Identify and justify a real dataset that exhibits similar structural challenges.
- Apply your modeling method to both datasets.
- Compare behavior between synthetic and real data.
- Deliver a short teaching presentation demonstrating the method live.

The focus is conceptual mastery, clarity of explanation, and reproducible implementation.

### Project Components

#### 1. Technical Report (PDF)

Must include:

- Explanation of the modeling method.
- Description of pathological synthetic dataset.

- Description and justification of real dataset.
- Mathematical formulation of the method.
- Experimental results and comparison.
- Discussion of limitations and practical implications.

## 2. GitHub Repository

Must include:

- Clean, organized, well-commented code.
- Instructions for reproducing results.
- Clear README explaining project structure.

## 3. Teaching Presentation (5–8 minutes)

You must:

- You must schedule a rehearsal time with MAJ Kuiper prior to the actual live-demo event.
- Clearly explain the method.
- Demonstrate results live in code.
- Show why the method matters.

## Grading Rubric (300 Points)

Criterion	Points	Score
<b>1. Conceptual Depth and Mathematical Understanding</b>	<b>75</b>	
– Clear explanation of method	30	_____
– Correct mathematical formulation	25	_____
– Insight into strengths and limitations	20	_____
<b>2. Experimental Design and Implementation</b>	<b>100</b>	
– Improved synthetic dataset experiment	35	_____
– Appropriate real dataset selection	25	_____
– Correct implementation and evaluation	40	_____
<b>3. Reproducibility and Code Quality (GitHub)</b>	<b>50</b>	
– Clean, well-structured repository	25	_____
– Clear instructions and documentation	25	_____
<b>4. Teaching Presentation</b>	<b>75</b>	
– Clarity of explanation	25	_____
– Effective live demonstration	25	_____
– Depth of insight and engagement	25	_____
<b>Total</b>	<b>300</b>	_____