

AI Club's Project Workshop



- Build your own ML project from scratch
- Add to your portfolio
- Compete for prizes

Project Officers



Noah Pragin
Project Manager

A senior majoring in CS with a custom option focusing on robotics and AI. Experience includes TAing for the machine learning course and working on autonomous vehicles.



Lauren Gliane
Project Officer

A 3rd year in Computer Science with a Data Science focus. Joined AI Club to learn AI and help make AI education and projects accessible and welcoming at OSU.



Kellen Sullivan
Project Officer

A senior majoring in CS with a custom AI focus and minoring in math. Experience includes completing multiple AI projects during a data engineering internship.



Ally Chen
Project Officer

A sophomore in CS with a custom focus in AI and Finance, with a Finance minor. Joined AI Club to gain early hands-on experience with AI, an opportunity often limited to upperclassmen.

What We're Building

Goal: Complete ML Project in 10 Weeks

- Work on a guided machine learning project of your choice. No prior experience necessary and we'll help you through every step!

What You'll Gain:

- **Portfolio Project:** A complete ML project to showcase
- **Hands-on Experience:** Real-world data science workflow
- **Resume Material:** Demonstrable AI/ML skills
- **Competition:** Chance to win prizes, AI Club merch and either compute credits or an Amazon gift card!
- **Community:** Work alongside other learners



10-Week Project Timeline Pt. 1

1

Setup and Tutorials

Set up your computer to work on your project and complete tutorials for new tools

2

Problem & Dataset Selection

Define problem, select dataset, identify potential models

3

Data Analysis & Preprocessing

Explore data, create training, validation, and test splits, feature engineering

4

Model Comparison & Tuning

Train 3+ models, basic hyperparameter tuning

10-Week Project Timeline Pt. 2

5

Catch-up / Get ahead

Catch up on anything you missed, polish your model, plan your deployment approach

6

Final Model Selection

Comprehensive tuning and test set evaluation on selected model

7

Model Analysis

Bias detection, interpretability, error analysis

8/9

Deployment
Implementation

Build your interface, integrate your model

10

Presentations & Winner
Selection

Present your projects and vote on a winner!

Team Structure & Logistics

- Team Size

- Work solo or in groups, whatever works best for you!
- Solo projects are absolutely doable and encouraged.
- Groups larger than 3 may have a hard time evenly splitting work

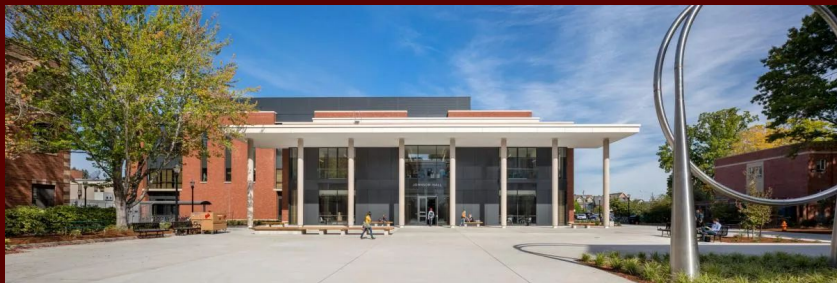
- Matchmaking

- Have friends? Form your own group!
- Flying solo? We'll help connect you with others!
- Want to work alone? Perfect!
- Scan the QR code and fill out the form to help us build teams



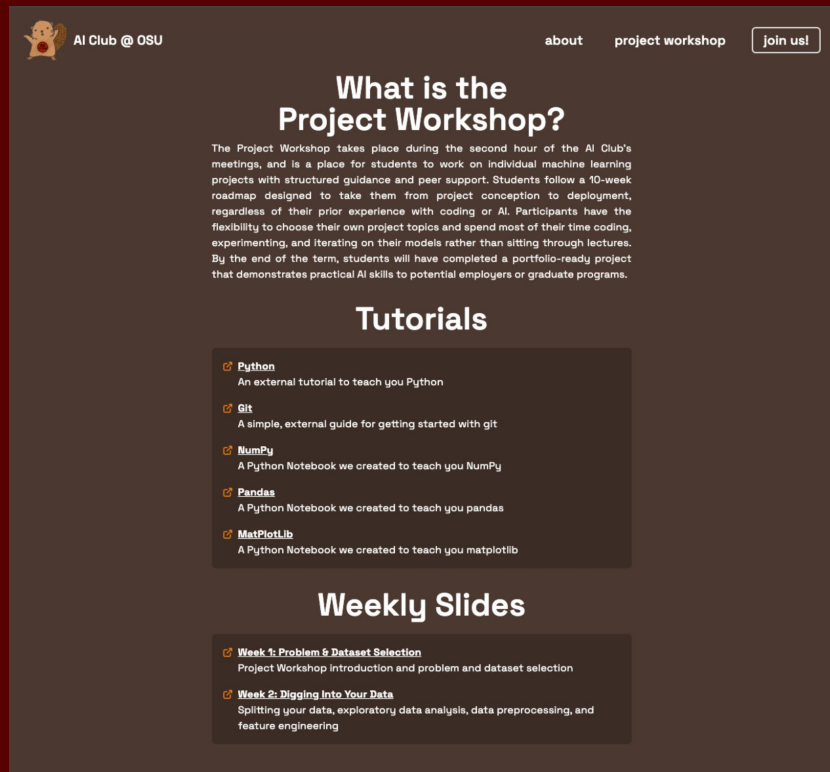
Support I

- During Meetings
 - We will walk around during meetings to check in and assist you
 - Recitation-style meetings, where we are happy to give you a hand if you ask
- Outside of Meetings
 - Noah's office hours: 10–11am T+Th in Johnson 121
 - Send questions to Discord!



Support II

- The Project Workshop page of the AI Club website has our slides, tutorials, and example project!
 - osu-ai.club/project-workshop
- The first and last slide will always have a QR code to our website



Packages / Tools

- NumPy
 - Numerical processing, used under the hood by everything else
- Pandas
 - Data organization and manipulation, used to aggregate your data before feeding it into a model
- Matplotlib
 - Plots and visualizations, used to visualize your dataset and model performance
- Scikit-learn
 - Machine learning models and data preprocessing, used for the training and evaluating models



Tips for Success

- Ask questions!
 - ML is a broad topic, and there is no way for us to cover everything
- Find something that interests you!
- Tell us if you see anything wrong or inconsistent!
- Consistency > Intensity



Feedback Survey

- Let us know how today's meeting went!



Week 1: Setting Up & Learning New Skills

- Definition of Done
 - Starting to think about what problem you want to work on
 - Access to ENGR servers
 - Code editor set up
 - Python ready to go
 - Tutorials complete for tools you're unfamiliar with
- Extra Credit
 - Problem identified and defined
 - Dataset selected and accessible



What Makes a Good Problem?

- Clean, available datasets
 - There is no model without data to learn from
- Interesting to you
- Some level of interactivity
- Classification problem?
 - Best suited to beginners, strongest support from us

| | A | B | C | D | E | F | G | H | I | J | K |
|----|-----|-----------|--------|------------|------------|----------|------------------------|----------|---------|---|---|
| 1 | Age | Sex | Job | Housing | Saving acc | Checking | Credit am | Duration | Purpose | | |
| 2 | 0 | 67 male | 2 own | NA | little | 1169 | 6 radio/TV | | | | |
| 3 | 1 | 22 female | 2 own | little | moderate | 5951 | 48 radio/TV | | | | |
| 4 | 2 | 49 male | 1 own | little | NA | 2096 | 12 education | | | | |
| 5 | 3 | 45 male | 2 free | little | little | 7682 | 42 furniture/equipment | | | | |
| 6 | 4 | 53 male | 2 free | little | little | 4870 | 24 car | | | | |
| 7 | 5 | 35 male | 1 free | NA | NA | 9055 | 36 education | | | | |
| 8 | 6 | 55 male | 2 own | quite rich | NA | 2855 | 24 furniture/equipment | | | | |
| 9 | 7 | 35 male | 3 rent | little | moderate | 6948 | 36 car | | | | |
| 10 | 8 | 61 male | 1 own | rich | NA | 8059 | 12 radio/TV | | | | |
| 11 | 9 | 28 male | 3 own | little | moderate | 5234 | 30 car | | | | |
| 12 | 10 | 25 female | 2 rent | little | moderate | 1295 | 12 car | | | | |
| 13 | 11 | 24 female | 2 rent | little | little | 4398 | 48 business | | | | |
| 14 | 12 | 22 female | 2 own | little | moderate | 1567 | 12 radio/TV | | | | |
| 15 | 13 | 60 male | 1 own | little | little | 1199 | 24 car | | | | |
| 16 | 14 | 28 female | 2 rent | little | little | 1403 | 15 car | | | | |
| 17 | 15 | 32 female | 1 own | moderate | little | 1262 | 24 radio/TV | | | | |
| 18 | 16 | 53 male | 2 own | NA | NA | 2424 | 24 radio/TV | | | | |
| 19 | 17 | 25 male | 2 own | NA | little | 8072 | 30 business | | | | |
| 20 | 18 | 44 female | 3 free | little | moderate | 12579 | 24 car | | | | |
| 21 | 19 | 21 male | 2 own | quite rich | NA | 3430 | 24 radio/TV | | | | |
| 22 | 20 | 48 male | 2 own | little | NA | 2134 | 9 car | | | | |
| 23 | 21 | 44 male | 2 rent | quite rich | little | 2647 | 6 radio/TV | | | | |
| 24 | 22 | 48 male | 1 rent | little | little | 2241 | 10 car | | | | |
| 25 | 23 | 44 male | 2 own | moderate | moderate | 1804 | 12 car | | | | |
| 26 | 24 | 26 male | 2 own | NA | NA | 2069 | 10 furniture/equipment | | | | |
| 27 | 25 | 36 male | 1 own | little | little | 1374 | 6 furniture/equipment | | | | |
| 28 | 26 | 39 male | 1 own | little | NA | 426 | 6 radio/TV | | | | |
| 29 | 27 | 42 female | 2 rent | rich | rich | 409 | 12 radio/TV | | | | |
| 30 | 28 | 34 male | 2 own | little | moderate | 2415 | 7 radio/TV | | | | |
| 31 | 29 | 63 male | 2 own | little | little | 6836 | 60 business | | | | |
| 32 | 30 | 36 male | 2 own | rich | moderate | 1913 | 18 business | | | | |
| 33 | 31 | 27 male | 2 own | little | little | 4020 | 24 furniture/equipment | | | | |
| 34 | 32 | 30 male | 2 own | moderate | moderate | 5466 | 18 car | | | | |
| 35 | 33 | 57 male | 1 rent | NA | NA | 1264 | 12 business | | | | |
| 36 | 34 | 33 female | 3 own | little | rich | 1474 | 12 furniture/equipment | | | | |



Limitations of Machine Learning

- Supervised ML Problem Types
 - **Classification:** Is this email spam or not spam?
 - **Regression:** What will the house price be?
- How to Frame a Problem for ML
 - Predict [specific outcome] based on [available data]
 - **Bad:** Detect COVID-19
 - **Good:** Predict COVID-19 severity based on lung CT images
- Guiding Questions
 - Is your prediction target specific?
 - Do you have data to learn from?
 - Can you quantitatively measure success?

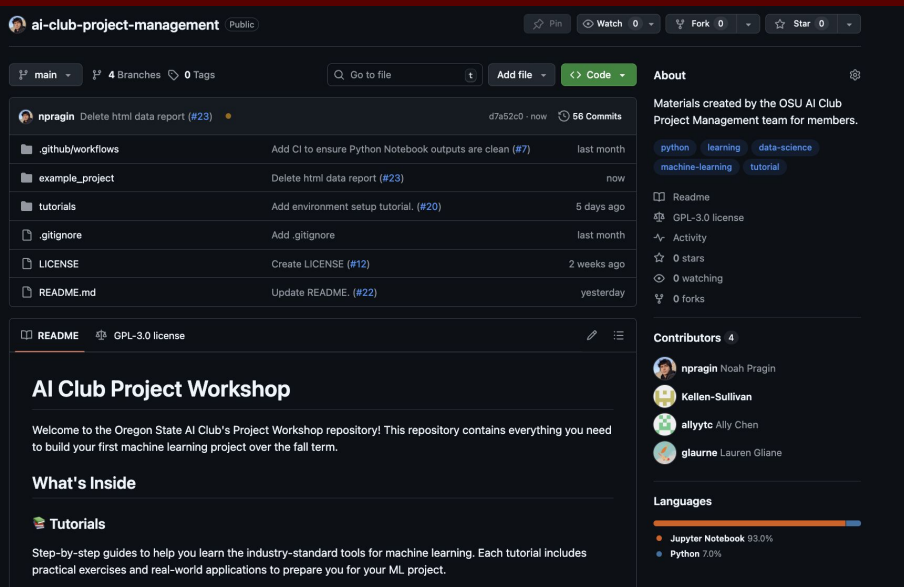
Project Ideas & Examples

- Pet Breed Classification
 - **Problem:** Identify cat and dog breeds from photos
 - **Dataset:** Oxford-IIT Pet Dataset
 - **Models:** CNN, ResNet
- Heart Disease Detection
 - **Problem:** Predict heart disease risk from medical data
 - **Dataset:** Cleveland Heart Disease Dataset
 - **Models:** Random Forest, Logistic Regression, SVM
- Network Security
 - **Problem:** Detect malicious network connections
 - **Dataset:** KDD Cup 99, NSL-KDD
 - **Models:** Decision Trees, Neural Networks, Ensemble



Tutorials (Ordered)

- Environment Setup
 - Skim this one even if you're set up, it calls out some good extensions for your code editor
- NumPy
- Pandas
- Scikit-learn
- Matplotlib?



Let's Build Something Amazing!

- Action Items
 - Start thinking about a problem to tackle!
 - Computer set up and ready to go
 - Complete tutorials on unfamiliar tools
- Resources
 - Tutorials on our website; scan the QR code!
 - Project idea slides
- Questions? Stuck?
 - Raise your hand! We're here to help

